STIC Search Report
Biotech-Chem Library

TO: Jon E Angell
Location: rem/2D20/2C18
Art Unit: 1635
Friday, July 01, 2005

Case Serial Number: 09/945173

From: Noble Jarrell
Location: Biotech-Chem Library
Rem 1B71
Phone: 272-2556

Noble.jarrell@uspto.gov

Search Notes

THIS PAGE BLANK (USPTO)

157842

Please rush. Thanks Chris

Chris Chan
TC 1600 New Hire Training Coordinator and SPE 1644
(571)-272-0841
Remsen, 3E89

-----Original Message-----
From: Angell, Jon E
Sent: Tuesday, June 28, 2005 5:19 PM
To: Chan, Christina
Subject: RUSH Sequence Search Request 09/945,173

Chrisitina,

I would like to request a RUSH search as indicated below. The cases may be allowable and I need the interference search ASAP.

If approved please forward the request to STIC and cc me.

Thanks,

Eric

SEARCH REQUEST FORM
Scientific and Technical Information Center

Examiner# : 78697
Art Unit : 1635
Phone Number: 571-272-0756
Date: 6/28/05
Serial Number: 09/945,173 (Meyers)
Mailbox & Bldg/Room Location: REMSEN 2C18
Results Format Preferred (circle): Paper

I would like to have a **standard and Interference** search performed using the following SEQ. ID NOs. from application : 09/945,173

SEQ ID NO. 1 (nucleic acid 1694 nucleotides in length)

****
STAFF USE ONLY

Searcher: **noble**
Searcher Phone: 2-
Date Searcher Picked up: 7/1/05
Date Completed:
Searcher Prep/Rev. Time: 5
Online Time: 3

****
Type of Search

NA#: 2   AA#: 1
Interference: ✓   SPDI:___
S/L:_____ Oligomer:_____
Encode/Transl:_____
Structure#:_____ Text:___
Inventor:____ Litigation:___

****
Vendors and cost where applicable
STN:_____
DIALOG:_____
QUESTEL/ORBIT:_____
LEXIS/NEXIS:_____
SEQUENCE SYSTEM: Complegu
WWW/Internet:_____
Other(Specify):_____

THIS PAGE BLANK (USPTO)

SEQ ID NO. 2 (protein 236 amino acids in length)
SQE ID NO: 3 (nucleic acid 711 nucleotides in length)

NOTE: SEQ ID NO: 1 is a full length DNA
          SEQ ID NO: 3 is the only the coding region of SEQ ID NO. 1
          SEQ ID NO: 2 is a protein encoded by SEQ ID NO: 1 and 3

Please feel free to contact me if you need any help...

Thanks for your help,

Jon Eric Angell
Art Unit 1635
Office: REMSEN 2D20
mailbox: REM 2C18
571-272-0756

THIS PAGE BLANK (USPTO)

```
                 GenCore version 5.1.6
          Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        June 30, 2005, 13:28:23 ; Search time 6313.94 Seconds
                                        (without alignments)
                                        10212.469 Million cell updates/sec

Title:         US-09-945-173-1
Perfect score: 1694
Sequence:      1 agggaaggcagtggcaagat.........aaacatgaattattgaacta 1694

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0
Searched:      3423954  seqs, 1903213470 residues

Total number of hits satisfying chosen parameters:   68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                 1: gb_est1:*
                 2: gb_est2:*
                 3: gb_htc:*
                 4: gb_est3:*
                 5: gb_est4:*
                 6: gb_est5:*
                 7: gb_est6:*
                 8: gb_gss1:*
                 9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                            SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1621.2	95.7	1867	3	CR619977	CR619977 full-leng
2	880	51.9	2013	3	AK028195	AK028195 Mus muscu
c 3	879	51.9	918	5	BX432920	BX432920 BX432920
4	868.6	51.3	2006	5	AK028269	AK028269 Mus muscu
5	799	47.2	912	5	BX348224	BX348224 BX348224
6	791.8	46.7	797	1	AL516309	AL516309 AL516309
7	791.8	46.7	918	3	BI862084	BI862084 603390970
8	770.4	45.5	1961	3	AK011280	AK011280 Mus muscu
9	750.6	44.3	815	5	BQ215951	BQ215951 AGENCOURT
c 10	750	44.3	875	7	CK456385	CK456385 917058 MA
11	748.2	44.2	950	5	BQ889351	BQ889351 AGENCOURT
12	747	44.1	855	7	CK455329	CK455329 915906 MA
13	718.4	42.4	720	7	CN361073	CN361073 328775509
14	718	42.4	718	5	BX099226	BX099226 BX099226
15	712.4	42.1	989	4	BM451532	BM451532 AGENCOURT
16	707.4	41.8	732	4	BI333788	BI333788 602999279
17	697	41.1	881	5	BU178665	BU178665 AGENCOURT
18	691.2	40.8	720	4	BG501635	BG501635 602548678
19	686	40.5	1118	4	AL517896	AL517896 AL517896
20	685.6	40.5	699	7	CN361072	CN361072 170006000
21	674	39.8	785	4	BI759699	BI759699 603045561
22	669.2	39.5	993	4	AK016099	AK016099 Mus muscu
23	664.2	39.2	902	4	BG258348	BG258348 602379905
24	662.8	39.1	959	4	BG031602	BG031602 602299747

```
                   636.4  37.6  25      4  BG025782  BG025782  602274775
             c     625.6  36.9  26      7  CN794203  CN794203  4129345 B
                   620.6  36.6  27      3  CA440480  CA440480  UI-H-ED0-
                   619    36.5  28      4  BI334986  BI334986  602998879
                   597.8  35.3  29      4  BG613092  BG613092  602640623
                   589.8  34.8  30      1  AA203636  AA203636  zx58b11.r
                   588.8  34.8  31      2  BF571796  BF571796  602078477
                   583    34.4  32      7  CN361075  CN361075  170004241
                   574.2  33.9  33      7  CF730316  CF730316  UI-M-GZ0-
                   567    33.5  34      4  BM674487  BM674487  UI-E-EJ0-
             c     564.4  33.3  35      1  AL516308  AL516308  AL516308
             c     562.6  33.2  36      4  BG080166  BG080166  H3050E03-
                   560.4  33.1  37      4  BM751558  BM751558  K-EST0027
                   560    33.1  38      2  BF106997  BF106997  601824352
                   559.6  33.0  39      2  BG400077  BG400077  602442152
                   556.8  32.9  40      2  BB610950  BB610950  BB610950
                   550.6  32.5  41      2  BB611463  BB611463  BB611463
                   549.8  32.5  42      6  CD512097  CD512097  AGENCOURT
             c     544.4  32.1  43      5  BQ013770  BQ013770  UI-1-BC1p
             c     540    31.9  44      1  AA662211  AA662211  ns66d03.r
             c     533.6  31.5  45      2  AW974777  AW974777  EST386867


                            ALIGNMENTS

RESULT 1
CR619977
LOCUS       CR619977               1867 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DA005YF10 of Neuroblastoma of Homo
            sapiens (human).
ACCESSION   CR619977
VERSION     CR619977.1  GI:50500784
KEYWORDS    HTC; CNSLT cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1867)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com URL ;
            Faraday Avenue
REFERENCE   2  (bases 1 to 1867)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
  COMMENT   1st strand cDNA was primed with a Noti-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1..1867
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DA005YF10"
                     /tissue_type="Neuroblastoma"
                     /plasmid="pCMVSPORT_6"
ORIGIN

Query Match             95.7%;  Score 1621.2;  DB 3;  Length 1867;
Best Local Similarity   99.8%;  Pred. No. 0;
Matches 1623;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy    62 CAGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTCGGGAAATC 121
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CAGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTCGGGAAATC 60
```

```
Qy  1202 TTTTCAAAGATGCAAAAATGGTGTTATTTAATTGTCTCCACCATTGTCACACAGGAAATG 1261
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 TTTTCAAAGATGCAAAAATGGTGTTATTTAATTGTCTCCACCATTGTCACACAGGAAATG 1200

Qy  1262 CCTAATAATAGCAACCCTTGTCTCCCTCTCTCCTCTCTCCTTTGCAAATGCTCAGTGACTGGA 1321
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 CCTAATAATAGCAACCCTTGTCTCCCTCTCTCCTCTCTCCTTTGCAAATGCTCAGTGACTGGA 1260

Qy  1322 AGAGGCGGACTAATAGCCAGAGTAAATATAAATAAATCAAAATTAATATACATAGAGAACAG 1381
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 AGAGGCGGACTAATAGCCAGAGTAAATATAAATAAATCAAAATTAATATACATAGAGAACAG 1320

Qy  1382 CAATACCAGAAAAAAAGAATTCTGGTAAAATGATGTGAAAAATTGACAGCTCCCTCACTC 1441
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 CAATACCAGAAAAAAAGAATTCTGGTAAAATGATGTGAAAAATTGACAGCTCCCTCACTC 1380

Qy  1442 TTAAGGTTGCTGCTATATACAGTCTAGGTTTTCTGTTGGAAATAGGTAGGGTAAAATCT 1501
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 TTAAGGTTGCTGCTATATACAGTCTAGGTTTTCTGTTGGAAATAGGTAGGGTAAAATCT 1440

Qy  1502 AAGACCTGCACAAGGGCAGTGAGAGACATTTACAGCCTCCTCTCCATTTGTTTTTTAAG 1561
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 AAGACCTGCACAAGGGCAGTGAGAGACATTTACAGCCTCCTCTCCATTTGTTTTTTAAG 1500

Qy  1562 GAAAAGTCAACTCCTGAAATGTCCCTTAGCTATAATCAGAAAACTAAGAAATATTATTCTG 1621
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 GAAAAGTCAACTCCTGAAATGTCCCTTAGCTATAATCAGAAAACTAAGAAATATTATTCTG 1560

Qy  1622 TGTCAACAATGTATTTATGGAGAGAAGTAAAAATAAGTTCCACAGCAACACATTTACATG 1681
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 TGTCAACAATGTATTTATGGAGAGAAGTAAAAATAAGTTCCACAGCAACACATTTACATG 1620

Qy  1682 AATTAT 1687
         ||||||
Db  1621 AATTAT 1626
```

RESULT 2
AK028195
LOCUS   AK028195   2013 bp   mRNA   linear   HTC 03-APR-2004
DEFINITION Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
            enriched library, clone:2610314P18 product:hypothetical RAS small
            GTPases, Rab subfamily/ATP/GTP-binding site motif A
            (P-loop)/Sigma-54 factor interaction domain containing protein,
            full insert sequence.
ACCESSION   AK028195
VERSION   AK028195.1 GI:26389871
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE   High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
   MEDLINE   99279253
    PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
   MEDLINE   20499374
    PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Suni,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,K., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,

```
Qy   122 CATCATGGACTGTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACC 181
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CATCATGGACTGTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACC 120

Qy   182 CAGAGAGAAGAAGACCTACTACATAGAATTATGATGGATGTTGGAGGCTCTGTGGGCAGTGCCA 241
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CAGAGAGAAGAAGACCTACTACATAGAATTATGATGGATGTTGGAGGCTCTGTGGGCAGTGCCA 180

Qy   242 GCAGGTGAAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTAC 301
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GCAGCGTGAAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTAC 240

Qy   302 ACGACTTAACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCA 361
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ACGACTTAACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCA 300

Qy   362 ACAGGGATTTGGTGCCAACTGGAGTCTTGGTGATAATGGGATTATGATCAAGAACAGT 421
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ACAGGGATTTGGTGCCAACTGGAGTCTTGGTGATAATGGGATTATGATCAAGAACAGT 360

Qy   422 TTGCTGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGACCAGATTCATGAAA 481
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TTGCTGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGACCAGATTCATGAAA 420

Qy   482 CAAAGCGCCATGAAGAGTTTTAACTAGGACTGCTTTCCTGCGACTGGATTTCAATCCAGAAG 541
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CAAAGCGCCATGAAGAGTTTTAACTAGGACTGCTTTCCTGCGACTGGATTTCAATCCAGAAG 480

Qy   542 AAATTAATTTGGACTGCACAAATCCACGTACTCCAGGTTCTTCCAATGCTGTCA 601
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 AAATTAATTTGGACTGCACAAATCCACGTACTCCAGGTTCTTCCAATGCTGTCA 540

Qy   602 AGCTCAGTAGGTTTTTGATAAGGTCATAGAGAGATACTTTTTAAGAGAAGTAATC 661
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 AGCTCAGTAGGTTTTTGATAAGGTCATAGAGAGAGATACTTTTTTAAGAGAAGTAATC 600

Qy   662 AGATTCCAGGCTTTCCTGATCGGAAAAGATTTGGGCCAGGAACATTAAAGAGCCTTCATT 721
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 AGATTCCAGGCTTTCCTGATCGGAAAAGATTTGGGCCAGGAACATTAAAGAGCCTTCATT 660

Qy   722 ATGACTGAATTACACTCATCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTTCACAGCT 781
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 ATGACTGAATTACACTCATCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTTCACAGCT 720

Qy   782 CATCTTGCTGTGTTCAATTATTACCATCACAGCCCTTTAACAAAATCATCTTAAAATGCT 841
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 CATCTTGCTGTGTTCAATTATTACCATCACAGCCCTTTAACAAAATCATCTTAAAATGCT 780

Qy   842 ACCCTTCAGCCTTACCCTTAATGAAAATGAAAGGAAGTGACAATACGGGAGTGACGAGGTCCAA 901
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 ACCCTTCAGCCTTACCCTTAATGAAAATGAAAGGAAGTGACAATACGGGAGTGACGAGGTCCAA 840

Qy   902 ACTTGTCCCTGTTCTCTGTGTTCCTTCGTCCCTGTTCCTGTGTATTATGTAAAA 961
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 ACTTGTCCCTGTTCTCTGTGTTCCTTCGTCCCTGTTCCTGTGTATTATGTAAAA 900

Qy   962 GCCTTGTGTAAAATGAGATGTGTTGTCAAAATGATGCAGTAAATGAGCAATGACACAGTGTAC 1021
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 GCCTTGTGTAAAATGAGATGTGTTGTCAAAATGATGCAGTAAATGAGCAATGACACAGTGTAC 960

Qy  1022 TGCAGAGAAAATTTACTCTTGCCTAGAACTGGAGGGTTTTTATGGGTCTGTAATTTTCCC 1081
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 TGCAGAGAAAATTTACTCTTGCCTAGAACTGGAGGGTTTTTATGGGTCTGTAATTTTCCC 1020

Qy  1082 ACACTCATTGCTGAAAGCTTAATTAAGTACTTCAAAAAGTATCTCCATTGTTTTACCTT 1141
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 ACACTCATTGCTGAAAGCTTAATTAAGTACTTCAAAAAGTATCTCCATTGTTTTACCTT 1080

Qy  1142 CTTGAGGGGAACGGTCTTGTTAACCAGCCCTGAGTTGTCTACCCCAAACAATCTCTGTCA 1201
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 CTTGAGGGGAACGGTCTTGTTAACCAGCCCTGAGTTGTCTACCCCAAACAATCTCTGTCA 1140
```

TITLE   Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE   20530913
PUBMED   11076861
REFERENCE   4
AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 2013)
AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT   cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.jp/.
            URL:http://fantom.gsc.riken.jp/.
FEATURES   Location/Qualifiers
     source   1..2013
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:2610314P18"
                     /db_xref="taxon:10090"
                     /clone="2610314P18"
                     /tissue_type="whole body"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="10 days embryo"
     CDS   16..726
                     /note="unnamed protein product; hypothetical RAS small
                     GTPases, Rab subfamily/ATP/GTP-binding site motif A
                     (P-loop)/Sigma-54 factor interaction domain containing
                     protein (InterPro|IPR003579, InterPro|IPR001687,
                     InterPro|IPR002078, evidence: InterPro)
                     putative"
                     /codon_start=1
                     /protein_id="BAC25804.1"
                     /db_xref="GI:26389872"
                     /translation="MASLDRVKVLVLGDSGVGKSSLVHLLCHNQVLGNPSWTVGCSVD
                     IRVHDYKRGTPEEKTYYIELWDVGGSVGSASSVKSTRAVFYNSVNGIILVHDLTNKKS
                     SQNLYRWSLEVLNRDAVPTGVLVTNGDYDREQFADNQIPLLVIGTKLDQIHETKRHEV
                     LIRTAFLAEDFNAEEINLDCTNPRPSAAGSNAVKLSRFFDKVIBKRYPFREGNQIPG
                     FSDRKRFGGGALKNFHCD"
     polyA_signal   1992..1997

                     /note="putative"
     polyA_site   2013
                     /note="putative"

ORIGIN
        Query Match   51.9%;   Score 880;   DB 3;   Length 2013;
        Best Local Similarity   79.1%;   Pred. No. 1.7e-219;
        Matches 1206;   Conservative 0;   Mismatches 280;   Indels 38;   Gaps 12;

Qy   11 GTGGCAAGATGGCGTCCCTGGATCGCGGTGAAGGTACTCGTGTGGGAGACTCAGGTGTTG 70
              ||||||||||||||||||||||||||| |||||||||||| ||||||||||||||||||
Db   8 GTGGCAAGATGGCGTCCCTGGATCGCGGTGAAGGTTCTGTGTGGGAGACTCAGGTGTTG 67

Qy   71 GGAAATCTTCGTTAGTCCATCCTATGCCAAAATCAAGTGCTCGGAAATCCATCATCATGGA 130
              ||||||||| || ||||||||||||||||| |||||||||||||||||| |||||||||||
Db   68 GGAAATCTTCACTCGTCCATCTCGTGCCACAATCAAGTGTTAGGAAATCCGTCATCGA 127

Qy   131 CTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAGA 190
              |||||||||||||||||||||||||| | |||||||||||||||||| ||||||||||||
Db   128 CTGTGGGCTGCTCGGTGGATATCAGAGTTCATGACTACAAAGAAGGGGACCCCTGAAGAGA 187

Qy   191 AGACCTACTACATAGAATTATGGGGATGTTGGAGGCTCTGTGTGGCAGTGCCAGCAGCGTGA 250
              || ||||| |||||||||| |||||| ||||||||||||||||||||||||||||||||||
Db   188 AGACATACTATATAGAACTATGGGGATGTTGGGAGTGCCAGCAGTGTGA 247

Qy   251 AAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATATTTCGTACACGACTAA 310
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   248 AAAGCACAAGAGCGGTGTTCTCACAACTCGTAAACGGCATCATTTAGTACATGACTAA 307

Qy   311 CAAATAGAAGTCCTCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATT 370
              |||||||||||||||| |||||| ||||||||||||||| ||||||||||||||||||
Db   308 CAAATAGAAGTCATCTCAAAACTTATATCGCTGGTCATTGGGAGTTCTCAATCGGGGATG 367

Qy   371 TGGTGCCAACTGGATTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGATA 430
              ||||||| |||||||||||||||||||| || |||||| ||||||||||||||||||||
Db   368 CAGTTCCAACCGAGTCCTGGTGACAAATGGGATTATGATACCGAGAACAGTTTGCTGATA 427

Qy   431 ACCAAATACCACCTGTTGTAATAGGGACTAAACTGGACTAAACTGAAACAAAGCGCC 490
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   428 ACCAAATCCCACTGTTGTAATAGGGACTAAACTGGACTAAACTCATGAAACCAAGCGCC 487

Qy   491 ATGAAGTTTAACTAGGACTGCTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATT 550
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   488 ATGAAGTTTAATCAGGACTGCTGCTTCCGTGGCTGAGGACTTCAATGCTGAAGAGATTAATT 547

Qy   551 TGGGACTGCACAAATCCACGGTACTTAGCTGACAGGTTCTTCCAATGCTGTCAAGCTCAGTA 610
              ||||| |||||||||| ||| ||||| |||||||||||| || ||||||||||||||||||
Db   548 TGGATTGCACAAACCCACGGCCCCTCAGCTGCAGGCTCCTCCAATGCTGTCAAGCTCAGTA 607

Qy   611 GGTTTTTGATAAGGTCATAGAAGAGATACTTTTAAGGAAGGTAATCAGATTCCAG 670
              || |||||||||||||||| |||| |||||| ||||||||||| |||||||||||
Db   608 GATTTTTGATAAGGTTATAGAAGAGATATTTTTTAGAGAAGGTAATCAGATTCCGG 667

Qy   671 GCTTTTCCTGATCGGAAAAGATTTGGCGGCAGGACATTAAAGAGCCTTCATTATGACTGAA 730
              ||||||||||||| |||||| ||||| ||||||||||| |||||||||||||||||||||
Db   668 GCTTTTCTTGATCGGCAAGAGGTTTGGAGAGGGCGCACTGAAGAGACTTCCACTGTGACTGAG 727

Qy   731 TTACACTCATCCTTTGGAAGAGTGAAGCAAGCAGTGGCCAGTTTTCACAGCTCATCTTGCT 790
              |||||||||| |||||| ||||||| |||||||||||| ||||| |||||||||||||||
Db   728 TGGCCCTCAGCC-CTGGAAGAGCAGGAACAGCAGCAGTGTCCACAGCTTCCTCTTGCT 786

Qy   791 GTGTTCAATTATTACCATCACAGCCTTTAACAAAATCATCTTAAATGTCACCCTTCAG 850
              |||||||||||||||| |||||||||||||||||| |||||| ||||||||||||||
Db   787 GTGTAGGGACATTAACTTCTCAGCC-TTTGGGGAAAAACATCTCAAAATCCTA------- 837

Qy   851 CCTTACCCTTTAATGGAAAAAATGAAAGGAAGTGACAATACGGAGGTCCAAACTTTGTCC 910
              ||||||| |||| |||||||| ||||||| |||||||||||| |||||||||||||||||
Db   838 CCTCACCCTCTCTTGCAAAATTGAAACGGAGGAAGCGCCATGAAGGTCCTAGCTCCGCCC 897

Qy   911 CTGTTCTCTGTGTCCTTTCCTTACCTTCGTCCCTGTGTATAGATTATGTAAAAGCCTTGTGT 970
              ||| | |||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   898 ATTCTCTGTA---TCCTCACCACCCTGTCCCTGATTATAGCTTATGTAATTAAAAGCCTTGTGA 954

```
Qy    971 AAATATGAGAGATGTGTCAAAATGATGCAGTAAATGAGCAATGACAGTGTACTGCAGAGAA 1030
          ||||||  |||| || ||||||||||||    ||||||||||||| || ||||||||||||
Db    955 CGACATAAAGGCTTGTCAAACCCATGCAGTAAAGAACTGGAGACA-AGGGCTGCAGCGAA 1013

Qy   1031 AATTTACTCTTGCCTAGAACTGGAGGGTTTTTATGGTCTGTAATTTCCCACACTCATT 1090
          |||||| ||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1014 GGCTGACTCTTCACTTGGAACCAGCCGG-TTTACAGGCCTGCGATTTTCCCACACTTGCT 1072

Qy   1091 GCTGAAAGCTTAATTAAGTACTTCAAAAAACGTATCCCATTGTTTGTTTACCTTCTTGAGGGG 1150
          ||||||||||||||||||||||||||||||||||||||||||||||  ||  || ||||||
Db   1073 GCTGAAAGCTTAATTAAGTACTTCAGAAACATATCGTGTTGTTTTA-CTTCCTGAGGGG 1131

Qy   1151 AACGGTCTTGTTAACCAGCCCTGAGTTGTCTACCCCAAACAATCTCGTCATTTTCAAAG 1210
          ||| ||||| ||||||  |||||||||||||||||||||||||||||||||||||||||
Db   1132 AAAGGTCATGTTAATCAGCCCTGAGTTATCGCCTGAGTTATCGCCTGCCATTTTAAAG 1191

Qy   1211 ATGCAAAATGGTGTTATTTAATTGTCTCCACCATTGTCACACACAGGAATGCCTAATAAT 1270
          |||||||||||||||||||||||||||||||| || ||||||||||||||||||||||||
Db   1192 AAGCCAAATGGTGTTATTTCATTTTCTCCATCCTCATTGTCACACACAGGATGCCTAAGAAT 1251

Qy   1271 AGCACCCCGTTGTCTCCCCTTTGCAAATGGCTCAGTGACTGG---AAGAGGC 1327
          |||||||||||||||||||||||||||||||||||||||||||
Db   1252 AGCAACCC--CATCTCCCTCGTCCTCTGGGAAGGGCTAGTGTCTGGGAAGAGCGAT 1309

Qy   1328 GGACTAATAGCCAGAGTTAAAATATAAATAAATACAAATTAATAATACATAGAGAACAGCAATAC 1387
          |||||||||||||||||||||  |||  || ||||||| |||| ||||||||||||||||||
Db   1310 GGACTAATAGCTGGAGTTAAAATAAAATAATTAATACATAGGCGAACAG----AC 1365

Qy   1388 CAGAAAAAAAGAATCTGTGAAAATGATGTGAAAAATTGACAGCTCCCTCACTCTTAAGG 1447
          ||||||||||||||||||         ||||||||||||||||||||||||||||||||
Db   1366 CAGAGAGAGGGT----------CTGTGAGACTCTAACAGCTCCTCTGTC-CAAGG 1412

Qy   1448 TTGCTGCTATATACAGTCCTAGGTTTTCTGTTTGGAAATAGTAGGTAGGTAAAATCTAAGACC 1507
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1413 TTGCAGCTGTCGGAAGGCTCTTTCTCGCTTTCGTTTGGAAAGTGGGTGAAAATCTGAGACC 1472

Qy   1508 TGCACAAGGCAGTGAGAGACATT 1531
          ||| || |||||||||||  |||
Db   1473 TGTACAAGGCCAGTAAGACAGATT 1496
```

RESULT 3

BX432920/c

LOCUS   BX432920   918 bp   mRNA   linear   EST 04-MAY-2004
DEFINITION BX432920 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
            CS0DA005YF10 3-PRIME, mRNA sequence.
ACCESSION   BX432920
VERSION   BX432920.2 GI:47002492
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 918)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE   Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT   On May 15, 2003 this sequence version replaced gi:30775154.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 1912.r,
            and it belongs to a clone representative of this cluster. For more
            information about this cluster and the virtual cDNA, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAI001ZA07_CS00049_2&c=1912.r
FEATURES   Location/Qualifiers

source   1..918
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CS0DA005YF10"
                /tissue_type="NEUROBLASTOMA"
                /clone_lib="Homo sapiens NEUROBLASTOMA"
                /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
                with a NotI-oligo(dT) prImer. Five prime end enriched,
                double-strand cDNA was digested with Not I and cloned into
                the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                Library was not normalized."

ORIGIN

Query Match   51.9%; Score 879; DB 5; Length 918;
Best Local Similarity   98.9%; Pred. No. 2.5e-219;
Matches 904; Conservative   0; Mismatches   8; Indels   2; Gaps   2;

```
Qy    328 CAAAACTTGCGTCGTTGGTCATTGGAAGCTCCAACAGGGATTTGGTGCCAACTGCGAGTC 387
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    918 CAAAACTTGCGTCGTTGGTCATTGGTCATTGGAAGCTCCACAGGGNATTTGGTGCAAACTGAAGTC 859

Qy    388 TTGGTGACAAATGGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATA-CCACTGTT 446
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db    858 TTGGTGACAAATGGGGATTATGATCAAGAACAGTNTGCTGATAACCAAATACCCACTGTT 799

Qy    447 GGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCCATGAAGTTTAACTAG 506
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    798 GGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCCATGAAGTTTAACTAG 739

Qy    507 GACTGCTTCCTGGCTGAGGATTCAATCCAATGTGTCAAGCTGTGTAATCAGATTCCAGGCTTCCTGATCGGAA 566
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    738 GACTGCTTCCTGGCTGAGGATTCAATCCAATGTGTCAAGCTGTGTAATCAGATTCCAGGCTTCCTGATCGGAA 679

Qy    567 ACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCGTCAGTAGGTTTTTGATAAGGT 626
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    678 ACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCGTCAGTAGGTTTTTGATAAGGT 619

Qy    627 CATAGAGAAGAAGAGATACTTTTAAGAGAAGAAGGTAATCAGATTCCAGGCTTCCTGATCGGAA 686
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    618 CATAGAGAAGAAGAGATACTTTTAAGAGAAGAAGGTAATCAGATTCCAGGCTTCCTGATCGGAA 559

Qy    687 AAGATTTGGGCAGGAACATTAAAGAGCCTTCATTATGACTGAATTACTACATCCTTTG 746
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    558 AAGATTTGGGCAGGAACATTAAAGAGCCTTCATTATGACTGAATTACTACATCCTTTG 499

Qy    747 GAAGAGTGAGCAAGCAGTGGCAGTTTTTCACAGCTCATCTTGCTGTGTTCAATTATTACC 806
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    498 GAAGAGTGAGCAAGCAGTGGCAGTTTTTCACAGCTCATCTTGCTGTGTTCAATTATTACC 439

Qy    807 ATCACAGCCCTTTAACAAAATCATCTTAAAATGCTACCCTTCAGCCTTACCCTTTAATGG 866
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    438 ATCACAGCCCTTTAACAAAATCATCTTAAAATGCTACCCTTCAGCCTTACCCTTTAATGG 379

Qy    867 AAAAATGAAAGGAAGTGACAATACGGGAGGTCCAAACTTGTCCCGTGTCTCTGTGTCC 926
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    378 AAAAATGAAAGGAAGTGACAATACGGGAGGTCCAAACTTGTCCCGTGTCTCTGTGTCC 319

Qy    927 TTACCTTCTGTCCCCGTGTATAGATTATGTAAAATATGAGATGTTGT 986
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    318 TTACCTTCTGTCCCCGTGTATAGATTATGTAAAATATGAGATGTTGT 259

Qy    987 CAAAATGATGCAGTAAATGAGCAATGACAGTGTACTGCAGAGAAAATTACTCTTGCCTA 1046
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    258 CAAAATGATGCAGTAAATGAGCAATGACAGTGTACTGCAGAGAAAATTACTCTTGCCTA 199

Qy   1047 GAACTGGAGGGTTTTTATGGTCTGTAATTTCCCACACTCATTGCTGAAAGCTTAATTA 1106
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    198 GAACTGGAGGGTTTTTATGGTCTGTAATTTCCCACACTCATTGCTGAAAGCTTAATTA 139

Qy   1107 AGTACTTCAAAAACGTATCCCATTGTTTTACCTTCTTGTTTACCTTCTTGTTTACC 1166
          ||||||||||||||||||||||||||||||||||||||||||||
Db    138 AGTACTTCAAAAACGTATCCCATTGTTTTACCTTCTTGTTTACC 79
```

```
Qy  1167 AGCCCTGGAGTT-GTCTACCCCAAACAATCTCGTCATTTTCAAGATGCAAAATGGTGTT 1225
         |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db    78 AGCCCTGGAGTTGGTCACCCCAAACAATCTCTGTCATTTTCAAGATGCAAAATGGTGTT 19

Qy  1226 ATTTAATTGTCTCC 1239
         |||||||||||
Db    18 ATTTAATTGTCTCC 5
```

RESULT 4
AK028269
LOCUS   AK028269   2006 bp   mRNA   linear   HTC 03-APR-2004
DEFINITION Mus musculus 13 days embryo head cDNA, RIKEN full-length enriched
            library, clone:3110079H15 product:hypothetical RAS small GTPases,
            Rab subfamily/ATP/GTP-binding site motif A (P-loop)/Sigma-54 factor
            interaction domain containing protein, full insert sequence.
ACCESSION   AK028269
VERSION   AK028269.1 GI:26080787
KEYWORDS   HTC; CAP trapper.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS   Carninci,P. and Hayashizaki,Y.
TITLE   High-efficiency full-length cDNA cloning
JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
MEDLINE   99279253
PUBMED   10049636
REFERENCE   2
AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE   Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE   20499374
PUBMED   11042159
REFERENCE   3
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,A., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE   RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE   20530913
PUBMED   11076861
REFERENCE   4
AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
TITLE   Functional annotation of a full-length mouse cDNA collection
JOURNAL   Nature 409, 685-690 (2001)
REFERENCE   5
AUTHORS   The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL   Nature 420, 563-573 (2002)
REFERENCE   6 (bases 1 to 2006)
AUTHORS   Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT   cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.jp/
            URL:http://fantom.gsc.riken.jp/.
FEATURES   Location/Qualifiers
     source   1..2006
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:3110079H15"
                     /clone="3110079H15"
                     /tissue_type="head"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="13 days embryo"
                     /db_xref="taxon:10090"
     misc_feature   1..2006
                     /note="hypothetical RAS small GTPases, Rab
                     subfamily/ATP/GTP-binding site motif A (P-loop)/Sigma-54
                     factor interaction domain containing protein
                     (InterPro|IPR003579, InterPro|IPR001687,
                     InterPro|IPR002078, evidence: InterPro)"
ORIGIN

Query Match   51.3%; Score 868.6; DB 3; Length 2006;
Best Local Similarity   79.1%; Pred. No. 1.6e-216;
Matches 1205; Conservative   0; Mismatches 279; Indels 39; Gaps 13;

Qy 13 GGCAAGATGGCGTCCCTGGATCGGTGAAGGTACTGGTTGTTGGGAGACTCAGGTGTTGGG 72
        |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   2 GGCAAGATGGCGTCCCTGGATCGGTGAAGGTTCTGGTGTTCGGGAGACTCAGGTGTTGGG 61

Qy 73 AAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACT 132
        |||||| ||||||||||||||| |||||| |||||||| ||||||||||||||||||||||
Db 62 AAATCTTCACTCGTCCATCTTCTGTGCCACAATCAAGTGTTAGGAAATCCGTCATGGACT 121

Qy 133 GTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAGAAG 192
        |||||||||||| |||||| ||||||||||| || |||| ||||| |||||| ||||||||
Db 122 GTGGGCTGCTCGGTGGATATCAGAGTTCATGACTACACAAGAAGGAGGACCCCTGAAGAGAAG 181

Qy 193 ACCTACTACATAGAATTATGGAGGTTGGAGGCTCTG-TGGGCAGTGGCAGCAGCGTGAA 251
        || |||||| |||||| ||||| |||||||||||||| |||||||||||||||| |||||
Db 182 ACATACTATATAGAACTATGGGAGGTTGGAGGCTCTGCTGGGCAGTGGCAGCAGTGTGAA 241

Qy 252 AAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGACTTAAC 311
        |||||||||||| |||| ||||||||| ||||| || || || | |||||| |||||||
Db 242 AAGCACAAGAGAGCGGTGTTCTACAACTCTGTAAACGGCATCATTTTAGTACATGACTTAAC 301

Qy 312 AAATAAGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTTCAACAGGGATT 371
        |||||||||||| || ||||||| | | |||||||||||||||| ||| ||| ||||| |
Db 302 AAATAAGAAGTCATCTCAAAACTTATATCGCTGGTCATTGGAAGTTCTCAATCGGGATGC 361

Qy 372 GGTGCCAACTGGCGAGTCTTGGTGACAAATGGTGGATTATCAAGAACAGTTTGCTGATAA 431
        ||| ||||| || ||||| ||| |||||||||| ||||| | |||||||||||||||||
Db 362 AGTTCCAACCGGAGTCCTGTGTGACAAATGGGGATTATGACCGAGAACAGTTTGCTGATAA 421

Qy 432 CCAAATACCACCTGTTGGTAATAGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCA 491
        ||||| | ||||||||||||||||| || ||||||||||||||||||||||| ||||||||
Db 422 CCAAATCCCACCTGTTGGTAATAGGGACCAAACTGGACCAGATTCATGAAACCAAGCGCCA 481
```

```
RESULT 5
BX348224

LOCUS       BX348224                912 bp    mRNA        linear   EST 23-APR-2004
DEFINITION  BX348224 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
            cDNA clone CS0DC020YI08 5-PRIME, mRNA sequence.
ACCESSION   BX348224
VERSION     BX348224.1  GI:30383253
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 912)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen. This sequence belongs to sequence cluster
            10682.r
            For more information about this cluster, see
            http://www.genoscope.cns.fr/cdna?s=CS0BAF007ZA11_AF00658_1&c=10682.
            r.
FEATURES             Location/Qualifiers
     source          1..912
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DC020YI08"
                     /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match             47.2%;  Score 799;  DB 5;  Length 912;
Best Local Similarity   95.1%;  Pred. No. 2.5e-198;
Matches  834;  Conservative   0;  Mismatches  42;  Indels  1;  Gaps  1;
```

(Sequence alignment follows, Qy = Query, Db = Database)

```
Qy  732 TACACTCATCCTTTGGAAGAGTGAGCAAGCAGTGAGCAAGCAGTTTTCACAGCTCATCTTGCTG 791
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 TACACTCATCCTTTGGAAGAGTGAGCAAGCAGTGAGCAAGCAGTTTTCACAGCTCATCTTGCTG 60

Qy  792 TGTTCAATTATTACCATCACAGCCTTTAACAAAATCATCTTAAAATGCAAAATCATCTTAAAATGCTACCCTTCAGC 851
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TGTTCAATTATTACCATCACAGCCTTTAACAAAATCATCTTAAAATGCAAAATCATCTTAAAATGCTACCCTTCAGC 120

Qy  852 CTTACCCTTTAATGGAAAAATGAAAAGGAAGTGACAATACGGAGGTCCAAACTTTGTCCC 911
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CTTACCCTTTAATGGAAAAATGAAAAGGAAGTGACAATACGGAGGTCCAAACTTTGTCCC 180

Qy  912 TGTTCTCTGTGTTCCTTACCTTTCTGTCCCTGTATAGATTATGTAGATTATGTAAAAGCCTTGGTA 971
            ||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  181 TGTTCTCTGTGTTCCTTACCTTTCTGTCCCTGTATAGATTATGTAGATTATGTAAAAGCCTTGGTA 240

Qy  972 AATATGAGATGTGTCAAAATGATGAGCAATGACAGTGTACTGACAGAAA 1031
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 AATATGAGATGTGTCAAAATGATGAGCAATGACAGTGTACTGACAGAAA 300

Qy 1032 ATTTACTCTTGCCTAGAACTGGAGGGTTTTATGGGTCTGTAATTTTCCCACACTCATTG 1091
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 ATTTACTCTTGCCTAGAACTGGAGGGTTTTATGGGTCTGTAATTCTCCCACACTCATTG 360

Qy 1092 CTGAAAGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTACCTTCTTGAGGGGA 1151
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

(Lower section — continuation of alignment)

```
Qy  492 TGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTCAATCCAGAGAAATTAATTT 551
            ||||||||||||||||||||||
Db  482 TGAAGTTTAATCAGGACTGCTTTCCTGGCTGAGGACTTCAATGCTGAAGAGATTAATTT 541

Qy  552 GGACTGCACAAATCCACGGTACTTAGCTGCAAATCGTGTCCAATGCTGTCAAGCTCAGTAG 611
            ||||||||||||||||||||||||||||
Db  542 GGATTGCACAAACCCACGGTCCTCAGCTGCGAGGCTCCTCCCAATGCATGTCAAGCTCAGTAG 601

Qy  612 GTTTTTGATAAGGTCATAGAGAGAGATACTTTTTAAGAAGTAATCAGATTCCAGG 671
            ||||||||||||||||||||||||
Db  602 ATTTTTTGATAAGGTTATAGAGAGAGATATTTTTTTAGAGAAGTAATCAGATTCCGGG 661

Qy  672 CTTTCCTGATCGGAAAAAGATTGGGGCAGGAACATTAAAGAGCCCTCATTATGACTGAAT 731
            |||||||||||||||||||
Db  662 CTTTTCGATCGGAGAGGTTGGAGGAGGGGGCACTGAGAAACTTCCACTGCTGTGACTGAGT 721

Qy  732 TACACTCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAAGCTGGCAGTTTTTCACAGCTCATCTTGCTG 791
            |||||||||||||||||||||||||||||
Db  722 GGCCCTCAGCC-CTGGAAGAGCAAGCAAGGGACAGCAGCAGCAGTCCCAGAGCTCCTCTTGCTG 780

Qy  792 TGTTCAATTATTACCATCACAGCCTTTAACAAAATCATCTAAAATGTCACCCTTCAGC 851
            ||||||||||||||||||||||||||||||
Db  781 TGTAGGGACATTAACTTCTCAGGC-TTTGGGAAAAACATCTCAAATGTCA--------C 831

Qy  852 CTTACCCTTTAATGAAAAATGAAAAGGAAGTGACAATACGGAGGTCCAAACTTGTCCC 911
            ||||||||||||||||||||
Db  832 CTCACCCCTCCTTGAAAAATTGAAAAGGAAGGCAGCGCCATGAGGTCCTAGCTCTGACC 891

Qy  912 TGTTCTCTGTGTTCCTTACCTTTCTGTCCCTGTATAGATTATGTAGATTATGTAAAAGCCTTGGTGTA 971
            |||||||||
Db  892 TTCTCTGTA---TCCTCACCACCCGTCCCCGATTATAGCTATGTTATGTAAAAGCCTTGGTGAC 948

Qy  972 AATATGAGATGTGTCAAAATGATGAGCAATGACAGTGTACTGACAGAGAAA 1031
            ||||||||| ||||||||||| | ||||||||||||||||||
Db  949 GACATAAAGGCTGTGTCAAACCCATGCAGTGCAGTAAAGAGACTGGAGACA-AGGGCTGCAGCGAAG 1007

Qy 1032 ATTTACTCTTGCCTAGAAACTGGAGGGTTTTATGGGTCTGTAATTTTTCCCACACTCATTG 1091
            ||||||||||||||||||||||||
Db 1008 GCTGACTCTTACTTGGAAACCAGCGGG-TTTTACAGGCCTGCGATTTTCCCACACTTGCTG 1066

Qy 1092 CTGAAAGCTTAATTAAGTACTTCAAAAAGTATCTCCATTGTTTTTACCTTCTTGAGGGGA 1151
            |||||||||||||||||||||||||||||||
Db 1067 CTGAAGGCTTAATTAAGTACTTCAGAAAACATATCTGTGTGTTTTTA-CTTCCTGAGGGGA 1125

Qy 1152 ACGGTCTCGTTAACCAGCCAGAGTAAAAATCTCTGTCATTTTCAAAGA 1211
            |||||||||||||||||||||||||||||||||||| |||||||
Db 1126 AAGGTCATGTTAATCAGCGGCCCGAGTTATCCACCCCAAACAACCATCCCTGCCATTTTAAAAGA 1185

Qy 1212 TGCAAAATGGTGTTATTTAATTGTCTTCCACCATTGTCACACAGGAATGCCTAATAATA 1271
            ||||||||| ||||||
Db 1186 AGCCCAAATGGTGTTATTTATTTCTTCCATCCTCATTACACACGGGGATGCCTAAGAATA 1245

Qy 1272 GCAACCCTTTGTCTCCCTCTCTCCTTTTGCAAATGGCTCAGTGACTGG---AAGAGGGCG 1328
            |||||||||||| ||||| ||||||||||
Db 1246 GCAACCC--CATCTCCTCTGCTCTCTCCTTGGGAAGGCGTTAGTGTCTGGGAAGAGCGATG 1303

Qy 1329 GACTAATAGCCAGAGTAAAATAATTAATACAAATTAATAATACATTAGAGAACAGCAATACC 1388
            |||||||||||||||||||||||
Db 1304 GACTAATAGCTGAGTAAATAATAGAATTAATACAAATTAATAATACATGGCCGAACAG----ACC 1359

Qy 1389 AGAAAAAGAATTCTGGTAAAATGATGTGAAAAATTGACAGCTCCCTCCACTCTTAAGGT 1448
            |||||||||||||||| |||||
Db 1360 AGAAGAGAGGGT----------CGTGAGACTCTAACAGCTCCTCCTGCTCTC-CAAGGT 1406

Qy 1449 TGCTGCTATATACAGTCTCAGGTTTCTCGTTTGGAAATAGGTAGGGTAAAATCTAAGACCT 1508
            |||| ||||||
Db 1407 TGCAGCTGTCGGAAGGCTCTTTGCTTTCTGTTTGGGTTTTGTGTTTGGAAAGTGGGTGAAATCTGAGACCT 1466

Qy 1509 GCACCAAGGGCAGTGAGAGACATT 1531
            ||||||||||| ||||
Db 1467 GTACAAGGGCCAGTAAGAACAGAGATT 1489
```

```
Db   361 CTGAAAGCTTAATTAAGTACTTCAAAAACGTATCCCATTGTTTTTACCTTCTTGAGGGGA 420

Qy  1152 ACGGTCTTGTTAACCAGCCCTGAGTTGTCTACCCCAAACAATCTCGTCATTTCAAAGA 1211
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ACGGTCTTGTTAACCAGCCCTGAGTTGTCTACCCCAAACAATCTCGTCATTTCAAAGA 480

Qy  1212 TGCAAAATGGTGTTATTTAATTGTCTCCACCATTGTCACACACAGGAATGCCTAATAATA 1271
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TGCAAAATGGTGTTATTTAATTGTCTCCACCATTGTCACACACAGGAATGCCTAATAATA 540

Qy  1272 GCAACCCTGTCTCCCTCTTCCCTTGCAAATGGCTCAGTGGACTGGAGAGGACGGGAC 1331
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GCAACCCTGTCTCCCTCTTCCCTTGCAAATGGCTCAGTGGACTGGAGAGGACGGGAC 600

Qy  1332 TAATAGCCAGAGTTAAATATAATATCAAATTAATATACATAGAGAACAGCAATACCAGA 1391
          |||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   601 TAATAGCCAGAGTTAAATATAACTACAAATTAATACTACAGACAGCACTACCAGA 660

Qy  1392 AAAAAGAAATTCTGGTAAAATGATGTGAAAAATTGACAGCTCCCTCACTCTTAAGGTTGC 1451
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 AAAAAGAAATTCTGGCAAAATGATGTGAAAAATTGACAGCTCCCTCACTCTTAAGGTTGC 720

Qy  1452 TGCTATATACAGTCAGTTTTCTTGTTTGGAAAATAGGTAGGGT-AAAATCAAGACCTGC 1510
          |||||||||||||||     |||||||||||||||||||||||| |||||||||||||
Db   721 TGCTATATACAGNCTAGCGCCCTCGTGTCTGAAATAGGTAGGGTNAAAAATCTAAGACCTGC 780

Qy  1511 ACAAGGCAGTGAGAGACATTTACAGCCTCCTCCTCTCTATTTGTTTTTTAAGGAAAAGTCA 1570
          |||||||||||||||||||||||||||||||||||| |||  |||||||
Db   781 ACAAGGCAGTGAGAGACATTTACAGNCTGCTTTCTATTGTGTGGCGTTTGAGGAAAAAGTC 840

Qy  1571 ACTCCTGAAATGTCCCTTAGCTATAATACAGAAAAACTA 1607
          |||||||||||||||||||||| |||||||
Db   841 ACTTCTGCAATGCCCCCTCCTCATCCTATAACACATAAAAACACCAA 877
```

RESULT 6
LOCUS   AL516309   797 bp   mRNA   linear   EST 23-MAR-2004
DEFINITION   AL516309 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
           CS0DA005YF10 5-PRIME, mRNA sequence.
ACCESSION   AL516309
VERSION   AL516309.3   GI:45552953
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 797)
AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE   Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   On Feb 13, 2001 this sequence version replaced gi:30490812.
           Contact: Genoscope
           Genoscope - Centre National de Sequencage
           2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
           Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr·
           1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
           end enriched, double-strand cDNA was digested with Not I and cloned
           into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
           was not normalized. Library was constructed by Life Technologies, a
           division of Invitrogen.
           This sequence belongs to sequence cluster 1912.r,
           and it belongs to a clone representative of this cluster. For more
           information about this cluster and the virtual cDNA, see
           http://www.genoscope.cns.fr/cdna?s=CS0DA005DC05QP1&c=1912.r.
FEATURES   Location/Qualifiers
    source   1..797
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="CS0DA005YF10"
                    /tissue_type="NEUROBLASTOMA"
                    /clone_lib="Homo sapiens NEUROBLASTOMA"

                    /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
                    with a NotI-oligo(dT) primer. Five prime end enriched,
                    double-strand cDNA was digested with Not I and cloned into
                    the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                    Library was not normalized."

ORIGIN

Query Match   46.7%;   Score 791.8;   DB 1;   Length 797;
Best Local Similarity   98.9%;   Pred. No. 1.9e-196;
Matches   788;   Conservative   8;   Mismatches   1;   Indels   0;   Gaps   0;

```
Qy    62 CAGGTGTTGGAAATCTTCGTTAGTCCATCTCCTATGCCAAATCAAGTGCTCGGGAAATC 121
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CAGGTGTTGGAAATCTTCGTTAGTCCATCTCCTATGCCAAATCAAGTGCTCGGGAAATC 60

Qy   122 CATCATGACTGTGGGCTGCTCAGTGTCAGAGTTCATGATTACAAAGAAGGAACCC 181
          |||||||||||||||||||||||||||||||||||||||||||||:|||:|||||
Db    61 CATCATGACTGTGGGCTGCTCAGTGATGTCAGAGTTCATGATTACMAAGAGGAACCC 120

Qy   182 CAGAAGAAGAAGACCTACTACACATAGAAATTATGGATGTTGGAGGCTCTGTGGCAGTGCCA 241
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CAGAAGAAGAAGACCTACTACACATAGAAATTATGGATGTTGGAGGCTCTGTGGGCAGTGCCA 180

Qy   242 GCAGCCGTGAAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTAC 301
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GCCAGCCGTGAAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTAC 240

Qy   302 ACGACTTAACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCA 361
          |||||||||||||||||||||||||||||||||:|||||:||||||||||||||||||
Db   241 ACGACTTAACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCYCTCA 300

Qy   362 ACAGGGGATTTGGTGCCAACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGAACAGT 421
          |||||||||||||||||||||||||||||||||||:|||||||:|||||||||||||
Db   301 ACAGGGGATTTGGTGCCAACTGGAGTCTTGGTGACAAATGGGGATTATGATCMAGAACAGT 360

Qy   422 TTGCTGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAA 481
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TTGCTGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAA 420

Qy   482 CAAAGGGCCATGAAGTTTTAACTGACACTGCTTTCCTGGCTGAGGGATTTCAAATCCAGAAG 541
          ||||||||||||||||||||||||||||:|||:||||||||||||||||||||||||||
Db   421 CAAAGGGCCATGAAGTTTTAACTGACAGTGCYTTCCTGGCTGAGGGATTTCAAATCCAGAAG 480

Qy   542 AAATTAATTTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTGCCAATGCTGTCA 601
          |:||||||||||||||||||||||||||||:||||||||||||:|||||||||||||
Db   481 AMATTAATTTGGACTGCACAAATCCACGGTACTTAGCTGCMGGTGCGTCA 540

Qy   602 AGCTCAGTAGGTTTTTGATAAGGTCATAGAGGTCATAGAGAGAGATACTTTTAAGAGAAGGTAATC 661
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 AGCTCAGTAGGTTTTTGATAAGGTCATAGAGGTCATAGAGAGAGATACTTTTAAGAGAAGGTAATC 600

Qy   662 AGATTCCAGGCTTCCTGATCGGAAAAGAATTGGGGCAGGACAATTAAAGAGCCTTCATT 721
          |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 AGATTCCAGGCTTCCTGATCGGAAAAGAATTGGGGCAGGACAATTAAAGAGCCTTCATT 660

Qy   722 ATGACTGAATTACACTCCTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTCACAGCT 781
          ||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db   661 ATGACTGAATTACACTCCTTGGAAGAGTGAGCVAGCAGTGGCAGTTTTCACAGCT 720

Qy   782 CATCTTGCTGTGTTCAATTATTACCATCACCAGCCTTTTAACAAAATCATCTTAAAATGCT 841
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 CATCTTGCTGTGTTCAATTATTACCATCACCAGCCTTTTAACAAAATCATCTTAAAATGCT 780

Qy   842 ACCCTTCAGCCTTACCC 858
          ||||||||||||||||
Db   781 ACCCTTCAGCCTTACCC 797
```

RESULT 7
BI862084
LOCUS   BI862084   918 bp   mRNA   linear   EST 10-OCT-2001
DEFINITION   603390970F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5399969 5',
           mRNA sequence.

```
ACCESSION   BI862084
VERSION     BI862084.1  GI:16002831
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 918)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12019 row: l column: 18
            High quality sequence stop: 833.
FEATURES             Location/Qualifiers
     source          1..918
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5399969"
                     /tissue_type="mammary adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC 87"
                     /note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1,383 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
ORIGIN

Query Match      46.7%;  Score 791.8;  DB 4;  Length 918;
Best Local Similarity  95.2%;  Pred. No. 1.9e-196;
Matches  860;  Conservative  0;  Mismatches  37;  Indels  6;  Gaps  4;

Qy       27 CCTGGATCGCGGTGAAGGTACTGGTGTTGGGGAGACTCAGGTGTTGGGAAATCTTCGTTAGT 86
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        7 CCTGGATCGCGGTGAAGGTACTGGTGTTGGGGAGACTCAGGTGTTGGGAAATCTTCGTTAGT 66

Qy       87 CCATCTCCTATGCCAAAATCAAGTGCTCGGGAAATCCATCATGGACTGTGGGCTGCTCAGT 146
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       67 CCATCTCCTATGCCAAAATCAAGTGCTCGGGAAATCCATCATGGACTGTGGGCTGCTCAGT 126

Qy      147 GGATGTCAGAGTTCATGATGATTACAAAGAAGAAGAACCCCAGAAGAGAAGAAGACTACTACATAGA 206
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      127 GGATGTCAGAGTTCATGATGATTACAAAGAAGAAGAACCCCAGAAGAGAAGAAGACTACTACATAGA 186

Qy      207 ATTATGGATGTTGGAGGCTCGTGTGGGCAGTGCCAGCAGCCTGAAAAGCACAAGAGCAGT 266
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      187 ATTATGGATGTTGGAGGCTCGTGTGGGCAGTGCCAGCAGCCTGAAAAGCACAAGAGCAGT 246

Qy      267 ATTCTACAACTCCGTAAATGTATATTTTCGTACACGACTTAACAAAATAAGAGAGTCCTC 326
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      247 ATTCTACAACTCCGTAAATGTATATTTTCGTACACGACTTAACAAAATAAGAGAGTCCTC 306

Qy      327 CCAAAACTTGCTCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCAACTGGAGT 386
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      307 CCAAAACTTGCTCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCAACTGGAGT 366

Qy      387 CTTGGTGACAAATGGGGATTATGATCAAGAACAGTTGCTGATAACCAAATACCACTGTT 446
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      367 CTTGGTGACAAATGGGGATTATGATCAAGAACAGTTGCTGATAACCAAATACCACTGTT 426

Qy      447 GGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTTTTAACTAG 506
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      427 GGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTTTTAACTAG 486

Qy      507 GACTGCTTTCTGGCTGAGGATTCAATCCAGAAGAAAATTAATTTGGACTGCACAAATCC 566
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      487 GACTGCTTTCTGGCTGAGGATTCAATCCAGAAGAAAATTAATTTGGACTGCACAAATCC 546

Qy      567 ACGGTACTTAGCTGCAGGTTCTTCCAATGTCTCAAGCTGTGTCAGCTAGGTTTTTGATAAGGT 626
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      547 ACGGTACTTAGCTGCAGGTTCTTCCAATGTCTCAAGCTGTGTCAGCTAGGTTTTTGATAAGGT 606

Qy      627 CATAGAGAAGAGATACTTTTTAAGAGAAGGTAA--TCAGATCCAGGCTTTCCTGATCGG 684
            ||||||||||||||||||||||||||||||||||  |||||||||||||||||||||||
Db      607 CATAGAGAAGAGATACTTTTTAAGAGAAGGTAAATCACGATTCCAGGCTTTCCTGATCGG 666

Qy      685 AAAGATTTGGGGAGGAACATTAAAGAGAGCCTTCATTATGACTGAATTACACTCATCCTT 744
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      667 AAAGATTTGGGGCAGGAACATTAAAGAGCCTTCATTATGACTGAATTACACTCATCCTT 726

Qy      745 TGGAAGAGTGAGC-AAGCAGTGGCAG-TGTTTTCACAGCTCATCTGCTGTGTC--AATT 800
            |||||||||||| |||||||||||| |||||||||||||||||||||||||||  ||||
Db      727 TGGAAGAGTGAGCAAAGCAGTGGCAGTTGTTTTTCACAGCTCATCTGCTGTGTCCAATTT 786

Qy      801 ATTACCATCACAGCCTTTTAACAAAATCATCTTAAAATGCTACCCTTCAGCCTTACCCTT 860
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      787 ATTACCATCACAGTCTTTTAACAAAATCATCTTAAAATGCTACCTTCAGCCTTACCTTT 846

Qy      861 TAATGGAAAAATGAAAGGAAGTGACAATACGGGAAGTCCAAACTTTGTCCCTGTTCTCTG 920
            |||||||||||||||||||||||||||||||||||||||||||  |||||||||||||
Db      847 ATGGCACAACTGAACGGGAAGTGACCATACGGAAGGTCCAAAATTTGTCCCCTGTTCCTCC 906

Qy      921 TGT 923
            |||
Db      907 TGT 909
```

RESULT 8
AK011280
```
LOCUS       AK011280                1961 bp    mRNA    linear    HTC 03-APR-2004
DEFINITION  Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
            enriched library, clone:2610002D21 product:hypothetical RAS small
            GTPases, Rab subfamily/ATP/GTP-binding site motif A
            (P-loop)/Sigma-54 factor interaction domain containing protein,
            full insert sequence.
ACCESSION   AK011280
VERSION     AK011280.1  GI:12847298
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1
REFERENCE
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
  PUBMED    10049636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
  PUBMED    11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
```

```
JOURNAL     Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE     20530913
PUBMED      11076861
REFERENCE   4
AUTHORS     The RIKEN Genome Exploration Research Group Phase II Team and the
            FANTOM Consortium.
TITLE       Functional annotation of a full-length mouse cDNA collection
JOURNAL     Nature 409, 685-690 (2001)
REFERENCE   5
AUTHORS     The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
TITLE       Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL     Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 1961)
AUTHORS     Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
            Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
            Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
            Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
            Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
            Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
            Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
            Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
            Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.jp,
            URL:http://genome.gsc.riken.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.jp/) for further
            details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5' GAGAGAGAGAAGAGATCCAAGAGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. Second
            strand cDNA was prepared with the primer adapter of sequence [5'
            GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1..1961
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:2610002D21"
                     /db_xref="taxon:10090"
                     /clone="2610002D21"
                     /tissue_type="whole body"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="10 days embryo"
     CDS             17..406
                     /note="unnamed protein product; hypothetical RAS small
                     GTPases, Rab subfamily/ATP/GTP-binding site motif A
                     (P-loop)/Sigma-54 factor interaction domain containing
                     protein (InterPro:IPR003579, InterPro:IPR001687,
                     InterPro:IPR002078, evidence: InterPro)
                     putative"
                     /codon_start=1
                     /protein_id="BAB27514.1"
                     /db_xref="GI:12847299"
                     /translation="MASLDRVKVLVLGDSGVGKSSLVHLLCHNVQLGNPSWTVGCSVD
                     IRVHDYKEGTPEEKTYYIELWDVGGSVGSASSVKSTRAVFYNSVNGIILVHDLTNKKS
                     SQNLYRWSLEVLNRDAVPTGVLVTNGD"
```

```
     polyA_signal    1940..1945
                     /note="putative"
     polyA_site      1961
                     /note="putative"
ORIGIN
            Query Match          45.5%;  Score 770.4;  DB 3;  Length 1961;
            Best Local Similarity  75.9%;  Pred. No. 9.9e-191;
            Matches 1157;  Conservative  0;  Mismatches 276;  Indels  91;  Gaps  13;

Qy     11  GTGGCAAGATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTG   70
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9  GTGGCAAGATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTG   68

Qy     71  GGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGA   130
           ||||||||||||||| ||||| |||||||||||||||||||||||||||||||||||||
Db     69  GGAAATCTTCACTCGTCCATCTTCTGTGCCACCAATCAAGTGTTAGGAAATCCGTCATGGA   128

Qy    131  CTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAGA   190
           |||||||| || ||||||||| |||||||||||||||  ||||||||||||||||||||
Db    129  CTGTGGGCTGCTCGGTGGATATCAGAGTTCATGATTCATAAAGAGGGACCCCTGAAGAGA   188

Qy    191  AGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGA   250
           ||||||||||| ||||| |||||||| ||||||||||||||||||||||||||||||||
Db    189  AGACATACTATATAGAACATATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGTGTGA   248

Qy    251  AAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGACTAA   310
           ||||| ||||||||||||||  |||||||||||||||||||||||||| |||||||||
Db    249  AAAGCACAAGAGCGGTGTTCTACAACTCTGTAAACGGCATCATTTTAGTACATGACTAA   308

Qy    311  CAAATAAGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATT   370
           ||||||||||||| |||||||||  ||||||  |||||| ||||||| ||||||||
Db    309  CAAATAAGAAGTCATCTCAAAACTTATATCGCTGGTCATTGGAAGTTTCAATCAGGGATG   368

Qy    371  TGGTGCCAACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGAACAGTTTGCTGATA   430
           ||||||||||||||||||||||||||||||||||||||
Db    369  CAGTTCCAACCGGAGTCTTGGTGGTGACAAATG---------------------------   398

Qy    431  ACCAAATACCACTGTGGTAATAGGACTAAACTGGACCAGATTCATGAAACAAAGCGCC   490
           |||||||||||||||||||||||||||||||
Db    399  -------------GGGACTAAACTGGACCAGATTCATGAAACAAAGCGCC   435

Qy    491  ATGAAGTTTAACTAGGACTGCTTCCTGGCTGAGACTTCAATCCAGAAGAAATTAATT   550
           ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    436  ATGAAGTTTAATCAGGACTGCTTCCTGGCTGAGACTTCAATCCAGAAGAAGATTAATT   495

Qy    551  TGGACTGCACAAATCCACGGTACTTAGCTGGCAGGTTCTTCCAAGTCGTGTCAAGCTCAGTA   610
           ||||||| |||||||||||| ||||||||| |||||||||| ||||||||||||||||
Db    496  TGGATTGCACAAACCCACGGTCCTCGAGCGTCCTGCAGCGTCGCAGCGTGTCAAGCTCAGTA   555

Qy    611  GGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAGGTAATCAGATTCCAG   670
           |||||||||||||||||||||||||||||| ||||| |||||||||||||||||||||
Db    556  GATTTTTGATAAGGTTATAGAGAAGAGATATTTTTTTTGAGAAGGTAATCAGATTCCGG   615

Qy    671  GCTTTCCTGATCGGAAAAGATTTGGGCAGGAACATTAAAGAGCCTTCATTATGACTGAA   730
           |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db    616  GCTTTTCTGATCGGAAGAGAGTTTGGGAGGGGGCACTGAAGAACTTCCACTGTGACTGAG   675

Qy    731  TTACACTCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTTCACAGCTCATCCTGCT   790
           | |||||||||||| |||||||||||||||||||||||||| |||||||||||||||||
Db    676  TGGCCCTCAGCC-CTGGAAGAGCCAAGCAGCAGTCTCCACAGCTTCCTCTTGCT   734

Qy    791  GTGTTCAATTATTACCATCACAGCCTTTTAACAAAATCATCTTAAAATGCTACCCTTCAG   850
           ||||| |||   ||||||||||||||| ||||||| |||||||| || |||||||||
Db    735  GTGTAGGGACATTAACTTCTCAGGC-TTGGGAAAAACATCTCAAAATGCTA-------   785

Qy    851  CCTTACCCTTTTAATGAAAAATGAAAAAGGAAGTGACAATACGGAGGTCCAAACTTTGTCC   910
           |||||||||||| ||||||||||| ||||||||||||||||||| || |||||||||||
Db    786  CCTCACCCTCCTTTGGAAAATTGAAACGAAGGAAGCAGCGGCGATGATGAAGTCCTAGCTCTGCCC   845

Qy    911  CTGTTCTCTGTGTTCCTTACCTTTCTGTCCCCTGTGTATAGATTGATTATGTAAAGCCTTGTGT   970
           |||                || |||||||||||||||||||||||||| ||||||||||||||
Db    846  ATTCTCTGTA-------TCCTCACCACCGTGTCCCTGATTATAGCTTATGTAAAGCCTTGTGA   902
```

```
                                          /db_xref="taxon:9606"
                                          /clone="IMAGE:6045318"
                                          /tissue_type="embryonal carcinoma, cell line"
                                          /lab_host="DH10B (phage-resistant)"
                                          /clone_lib="NIH_MGC_92"
                                          /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
                                          Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                                          Average insert size 2.5 kb. Library enriched for
                                          full-length clones and constructed by Life Technologies.
                                          Note: this is a NIH_MGC Library."

ORIGIN

        Query Match      44.3%;  Score 750.6; DB 5;  Length 815;
        Best Local Similarity  98.4%;  Pred. No. 1.2e-185;
        Matches  779;  Conservative  0;  Mismatches  10;  Indels  3;  Gaps  2;

Qy      51 GTTGGGAGACTCAGGTGTTGGGAAATCTTCGTAGTCCATCTCCTAGTCCATGCCAAAATCAAGT 110
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GTTGGGAGACTCAGGTGTTGGGAAATCTTCGTAGTCCATCTCCTAGTCCATGCCAAAATCAAGT 60

Qy     111 GCTGGGAAATCCATCATGGACTGTGGGCTGCTCAGTGGATGGATGTCAGAGTTCATGATTACAA 170
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GCTGGGAAATCCATCATGGACTGTGGGCTGCTCAGTGGATGGATGTCAGAGTTCATGATTACAA 120

Qy     171 AGAAGGAACCCCAGAAGAGAAGACCTACTACATGAGAATTATGGGATGTTGGGAGGCTCTGT 230
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 AGAAGGAACCCCAGAAGAGAAGACCTACTACATGAGAATTATGGGATGTTGGGAGGCTCTGT 180

Qy     231 GGGCAGTGCCAGCAGCGTGAAAAGCACAAGAGCACAGCAGTATTCCACAACTCCGTAAATGTAT 290
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GGGCAGTGCCAGCAGCGTGAAAAGCACAAGAGCACAGCAGTATTCCACAACTCCGTAAATGTAT 240

Qy     291 TATTTTCGTACACGACTTAACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATT 350
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TATTTTCGTACACGACTTAACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATT 300

Qy     351 GGAAGCTCTCAACAGGGATTGGTGCCAACTGTTGGTGACAAATGGGGATTATGA 410
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GGAAGCTCTCAACAGGGATTGGTGCCAACTGTTGGTGACAAATGGGGATTATGA 360

Qy     411 TCAAGAACAGTTTGCTGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCA 470
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TCAAGAACAGTTTGCTGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCA 420

Qy     471 GATTCATGAAACAAAGCGCCATGAGAGTTTAACTAGGACTGCTTCCTGGCTGCTGAGGATTT 530
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GATTCATGAAACAAAGCGCCATGAGAGTTTAACTAGGACTGCTTCCTGGCTGCTGAGGATTT 480

Qy     531 CAATCCAGAGAAATTAATTGGACTGCACAAATCCACGTACTTAGCTGCAGGTTCTTC 590
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CAATCCAGAGAAATTAATTGGACTGCACAAATCCACGTACTTAGCTGCAGGTTCTTC 540

Qy     591 CAATGCTGTCCAAGCTCAGTAGGTTTTTGATAAGGTCATAGAAGAAGAGATACTTTTTAAG 650
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CAATGCTGTCCAAGCTCAGTAGGTTTTTGATAAGGTCATAGAAGAAGAGATACTTTTTAAG 600

Qy     651 AGAAGTAATCAGATTCCAGGCTTCCTGATCGGAAAAGATTTGGGGCAGGAACATTAAA 710
           |||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||
Db     601 AGAAGTAATCAGATTCCAGGCTTCCTGATNCGGAAAGATTTGGGGCAGGAACATTAAA 660

Qy     711 GAGCCTTCATTATGACTGAATTACACTCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAGT 770
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GAGCCTTCATTATGACTGAATTACACTCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAGT 720

Qy     771 TTTTCACAGTCATCTTGCT-GTGTTCAATTATTACCATCAC--AGCCTTTTAACAAAAT 827
           ||||||||||||||||||| ||||||||||||||||||||||  ||||||||||||||||
Db     721 TTTTCACAGTCATCTTGCTGGGGTTCAATTATTACCATCACCAGCCTTTTAACCAAGAT 780

Qy     828 CATCTTAAAATG 839
           ||||||||||||
Db     781 CATCTTAAAATG 792
```

```
Qy     971 AAATATGAGATGTTGTCAAAATGATGACGCAATGAGCAATGACAGTGTACTGCAGAGAA 1030
           ||||||||||||||||||||||  || |||| ||||||| ||||||||||||||||||||
Db     903 CGACATAAAGGCTTGTCAAACCCATGCAGTAAAGAGAACTGGAGACA-AGGGCTGCAGCGAA 961

Qy    1031 AATTTACTCTTGCCTAGAACTGGAGGGTTTTATGGGTCGTGTAATTTCCCACACTCATT 1090
           |||||||| ||| || |||| |||||||||||||||||||||||||||||||||||||||
Db     962 GGCTGACTCTTACTTGGACCAGCGGG-TTTTACAGGCCTGGCGATTTTCCCACACTTGCT 1020

Qy    1091 GCTGAAAGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTTTACCTTCTTGAGGGG 1150
           |||||||||||||||||||||||||||||||||||||||||| |||| ||||||||||||||
Db    1021 GCTGAAGGCTTAATTAAGTACTTCAGAAACATATCTGTGTGTTTTA-CTTCCTGAGGGG 1079

Qy    1151 AACGGTCTTGTTAACCAGCCTGAGTTGTCTACCCCAAACAATCTCTGTCATTTTCAAAG 1210
           ||||||||||||||||||||||||  ||  || ||| ||||| |||||||||||||  ||
Db    1080 AAAGTCATGTTAATCAGCCCTGAGTTATCTCACCCCAAACCATCCCTGCCATTTTAAAAG 1139

Qy    1211 ATGCAAAATGGTGTTATTTAATTGTCTCCACCATTGTCACACACAGGAATGCCTAATAAT 1270
           || ||||||||| |||||||| ||  || || |||||||| ||||||||  | |||||||
Db    1140 AAGCCAAATGGTGTTATTTCATTTTCTCCATCCTCATTACACACAGGATGCCTAAGAAT 1199

Qy    1271 AGCAACCCTTGTCTCTCCCTCTTCCTTTGCAAATGGCTCAGTGACTGG---AAGAGGC 1327
           ||||||| || |||||||||||||||| || |||| |||| ||||||| ||| |||||
Db    1200 AGCAACCC--CATCTCCCTCTGCTCTCGGAAAGGGCTTAGTGTCTGGGAAGAGCGAT 1257

Qy    1328 GGACTAAATAGCCAGAGTTAAAATATAAATACAAATTAATAATACATAGAGAACAGCAATAC 1387
           ||||||||||||||||||||||| ||| ||||| || |||||||| ||||||||||||||||
Db    1258 GGACTAAATAGCTGGAGTTAAAATAGAAATACAAATTAATAATACATAGGGCGAACAG----AC 1313

Qy    1388 CAGAAAAAAAGAATTCTGGTAAAATGATGTGAAAAAATTGACAGCTCCCTCCTCACTCTTAAGG 1447
           |||||       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1314 CAGAGAAGAGAAGGGT--------CTGTGAGACTCTAACAGCTCCCTCTGCTC-CAAGG 1360

Qy    1448 TTGCTGCTATATACAGTCTAGGTTTTCTGTTTGGAAATAGGTAGGTAAAATCTAAGACC 1507
           ||||||| ||| |||| ||||||| ||||| || ||||||||  ||||||||||||||||
Db    1361 TTGGCAGCTGTCGGAAGGCTCTTCGTCTTTCTGTTTGGAAAGTGGGTGAAAATCTGAGACC 1420

Qy    1508 TGCACAAGGGCAGTGAGAGACATT 1531
           |||||||||||||||||||||||
Db    1421 TGTACAAGGGCCAGTAAGACAGATT 1444
```

RESULT 9
BQ215951

LOCUS   BQ215951   815 bp   mRNA   linear   EST 02-MAY-2002
DEFINITION AGENCOURT_7558711 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6045318
            5', mRNA sequence.
ACCESSION   BQ215951
VERSION   BQ215951.1 GI:20397351
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 815)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13289 row: f column: 07
            High quality sequence stop: 614.
FEATURES   Location/Qualifiers
     source   1..815
                     /organism="Homo sapiens"
                     /mol_type="mRNA"

RESULT 10
CK456385/c
LOCUS   CK456385   875 bp   mRNA   linear   EST 12-JAN-2004
DEFINITION 917058 MARC 4PIG Sus scrofa cDNA 3', mRNA sequence.
ACCESSION   CK456385
VERSION   CK456385.1 GI:40803599
KEYWORDS   EST.
SOURCE   Sus scrofa (pig)
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1 (bases 1 to 875)
AUTHORS   Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A.,
            Nonneman,D.J., Wray,J.E. and Keele,J.W.
TITLE   Porcine EST collection using a normalized library constructed from
            embryos representing early developmental stages
JOURNAL   Unpublished (2003)
COMMENT   Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called with phred v0.020425.c and
            trimmed with the aid of the trim_alt option. Vector identified with
            cross_match v0.990329.
            Plate: TMW8028 row: B column: 12
            Seq primer: TAGAGGCACAGTCGAGG.
FEATURES   Location/Qualifiers
     source   1..875
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /tissue_type="pooled"
                     /clone_lib="MARC 4PIG"
                     /lab_host="DH10B"
                     /note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI;
                     Library made with combined RNA from day-10, day-13,
                     day-15, day-25, and day-30 whole embryos."
ORIGIN

Query Match   44.3%; Score 750; DB 7; Length 875;
Best Local Similarity 91.9%; Pred. No. 1.8e-185;
Matches   803; Conservative   0; Mismatches   70; Indels   1; Gaps   1;

Qy   101 AAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGGCTGCTCAGTGGATGTCAGAGTTC 160
          |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db   875 AAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGGCTGCTCAGTGGATGTCAGAGTTC 816

Qy   161 ATGATTACAAAGAAGAAGAACCCCAGAAGAGAAGAGACCTACTACATAGAATTATGGGATGTTG 220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   815 ATGACTACAAAGAAGAAGAACCCCAGAAGAGAAGAGACCTACTATATAGAATTATGGGATGTTG 756

Qy   221 GAGGCTCGTGTGGGCAGTGCCAGCAGCGGTGAAAAGCAAGAGCAGTATTCTCACAACTCCG 280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   755 GAGGCTCGGTGGGCAGTGCCAGCAGTGTGAAAAAGCACAAGAGCAGTATTCTCACAACTCTG 696

Qy   281 TAAATGGTATTATTTTCGTACACGACTTAACAAATAAGAAGTCCTCCCAAAACTTGCGTC 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   695 TAAATGGTATTATTTTAGTACATGACTTAACAAATAAGAAGTCATCCCAAAATTTGTATC 636

Qy   341 GTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCAACTGGAGTCTTGGTGACAAATG 400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   635 GTTGGTCATTGGAAGCTCTCAACAGGGATCTGGTGCCAACTGGAGTCTTGGTAACAAATG 576

Qy   401 GGGATTATGATCAAGAACAGTTGCTGATAACCAAATACCACTGTGGTAATAGGGACTA 460
          |||||||||||||||||| ||||||||||||||||||||| ||||||||||||||||||
Db   575 GGGATTATGACCGGGAACAGTTTGCTGATAACCAAATCCCACTGTGGTAATAGGGACTA 516

Qy   461 AACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCTGG 520
          |||||||||||||||||| ||||||||||||||||||||| |||||||||||||||||||
Db   515 AACTGGATCAGATTCATGAAACAAAACCAAGCGCCATGAAGTTTTAACTAGGACTGCTTTCCTGG 456

Qy   521 CTGAGGATTTCAATCAGAAGAAAATAATTTGGACTGCACAAATCCACGGTACTTAGCTG 580
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   455 CTGAGGATTTCAATGCAGAAGAAGATTAATTTGGATTGCACAAATCCACGGTACTTAGCTG 396

Qy   581 CAGGTTCTTCCAATGCTGTCAAGCTCAGTAGGTTTTGATAAGGTCATAGAAGAAGAGAT 640
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   395 CAGGTTCTTCCAATGCTGTCAAGCTCAGTAGGTTTTTTGATAAGGTCATAGAAGAAGAGAT 336

Qy   641 ACTTTTAAGAGAAGGTAATACAGATTCCAGGCTTCCTGATCGGAAAAGATTTGGGCAG 700
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   335 ACTTTTAAGAGAGATGGTAATACAGATTCCAGGCTTCCTGATCGGAAAAAGGTTTGGAGGAG 276

Qy   701 GAACATTAAAGAGCCTTCATTATGACTGAATTACACTTCATCTTTGGAGAGTGAGCAAG 760
          |||||||||||||||||||||||||| |||||| ||||| |||||||||||||||||||
Db   275 GAACATTAAAGAGCTTACATTATGACTGAATTACACTCATCCTTTGGAGAGGAGCAAT 216

Qy   761 CAGTGGCAGTTTTCACAGC-TCATCTTGCTGTGTCAATTATTACCATCACACAGCCTTTT 819
          ||||||||||||||||||| |||||| ||||| |||||||||||||| |||||||||||||
Db   215 TGGTGACAGTTTTTCACAGCGTTCTCCTGCTGTTTCTGTTTTGATTATTCCGTCCCCAGCCTTTT 156

Qy   820 AACAAAATCATCTTAAAAATGCTACCCTTCAGCCTTAACCCTTTAATGGAAAATGAAAGGA 879
          ||||||||||||||||||||||||||| |||||||| ||||||||||||||||||||||||
Db   155 AGCAAAATCATCTAAAAATGCTACCCTTCAGCCTCGCCGTTTAATGAAAACTGAAAGAA 96

Qy   880 AGTGACAATACGGGAGGTCCAAACTTTGTCCCGTGTTCTCTGTGTTCCTTACCTTTCTGTC 939
          ||||||||||||||||||||||| ||||||||||||||||||||||| ||||||||||||||
Db   95 AGTGACAGCGTGGGAGGTCCAAACTTTGTCCCCATTCTTTGTCCCGTGTTCTCTGTGTTCCTCACCTTTTTGTC 36

Qy   940 CCTGTGTATAGATTATCGTAAAAAGCCTTGTGTAA 973
          |||||||| |||| || |||||||||||||| |||
Db   35 CCTTTTTTATAGATTATATAAAAAGCCTTGGGAAAA 2

RESULT 11
BQ889351
LOCUS   BQ889351   950 bp   mRNA   linear   EST 16-AUG-2002
DEFINITION AGENCOURT 8113408 Lupski_dorsal_root_ganglion Homo sapiens cDNA
            clone IMAGE:6179062 5', mRNA sequence.
ACCESSION   BQ889351
VERSION   BQ889351.1 GI:22281365
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 950)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr James R. Lupski
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13559 row: j column: 23
            High quality sequence start: 19
            High quality sequence stop: 639.
FEATURES   Location/Qualifiers
     source   1..950
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6179062"
                     /sex="male"
                     /tissue_type="dorsal root ganglia"
                     /dev_stage="adult, 36 yr"
                     /lab_host="DH10B"
                     /clone_lib="Lupski dorsal root ganglion"
                     /note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:

```
CK455329
LOCUS           CK455329              855 bp    mRNA    linear    EST 12-JAN-2004
DEFINITION      915906 MARC 4PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION       CK455329
VERSION         CK455329.1  GI:40802543
KEYWORDS        EST.
SOURCE          Sus scrofa (pig)
  ORGANISM      Sus scrofa
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE       1 (bases 1 to 855)
  AUTHORS       Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A.,
                Nonneman,D.J., Wray,J.E. and Keele,J.W.
  TITLE         Porcine EST collection using a normalized library constructed from
                embryos representing early developmental stages
  JOURNAL       Unpublished (2003)
  COMMENT       Contact: Smith TPL
                USDA, ARS, US Meat Animal Research Center
                PO Box 166, Clay Center, NE 68933-0166, USA
                Tel: 402 762 4366
                Fax: 402 762 4390
                Email: smith@email.marc.usda.gov
                Single pass sequencing. Bases called with phred v0.020425.c and
                trimmed with the aid of the trim_alt option. Vector identified with
                cross_match v0.990329.
                Plate: TMW8028  row: B  column: 12
                Seq primer: GTAATACGACTCACTATAGGG.
FEATURES             Location/Qualifiers
     source          1..855
                     /organism="Sus scrofa"
                     /mol_type="mRNA"
                     /db_xref="taxon:9823"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /clone_lib="MARC 4PIG"
                     /note="Vector: pcDNA3.1; Site_1: EcoRI; Site_2: NotI;
                     Library made with combined RNA from day-10, day-13,
                     day-15, day-25, and day-30 whole embryos."
ORIGIN
```

Query Match   44.1%;   Score 747; DB 7; Length 855;
Best Local Similarity 92.9%; Pred. No. 1.1e-184;
Matches 794; Conservative 0; Mismatches 60; Indels 1; Gaps 1;

```
Qy    24 GTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTT 83
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCACT 60

Qy    84 AGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCAAGTGCTGTGGGCTGCTC 143
         ||| ||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db    61 AGTTCATCTTCTATGCCAAAATCAAGTGCTGGGAAATCCATCAAGTGCTGTGGGCTGCTC 120

Qy   144 AGTGGATGTCAGAGTTCATGATTACAAAGAAGAGGAACCCCAGAAGAGAAGAGACTACTACAT 203
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGTAGATGTCAGAGTTCATGATTACAAAGAAGAGGAACCCCAGAAGAGAAGAGACTACTATAT 180

Qy   204 AGAATTATGGGATGTTGGAGGCTTCGTGGGCGCTGCCAGCGCGTGAAAAGCACACAGAGC 263
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AGAATTATGGGATGTTGGAGGCTTCGTGGGCGCTGCCAGCGCGTGAAAAGCACACAGAGC 240

Qy   264 AGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGACTTAACACAATAAGAGAGTC 323
         ||| |||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   241 AGTATTCTACAACTCTGTAAATGGTATTATTTAGTACATGACTTAACACAATAAGAGAGTC 300

Qy   324 CTCCCAAAACTTGCCTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCAACTGG 383
         ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   301 ATCCCAAAATTGTATCGTTGGTCATTGTTGCATTGGAAGCTCTCAACAGGGATCTGGTGCCAACTGG 360

Qy   384 AGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGATAACCAACAATACCACT 443
         ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   361 AGTCTTGGTAACAAATGGGATTATGACCGGGAACAGTTTGCTGATAACCAAATCCCCACT 420

Qy   444 GTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTTTTAAC 503
```

Query Match   44.2%;   Score 748.2; DB 5; Length 950;
Best Local Similarity 97.3%; Pred. No. 5.5e-185;
Matches 793; Conservative 0; Mismatches 18; Indels 4; Gaps 3;

```
Qy   876 AGGAAGTGACAATACGGGAGGTCCAAACTTGTCCCGTGTTCTGTGTTCCTTACCTTTC 935
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    24 ATGAAGTGACAATACGGGAGGTCCAAACTTGTCCCGTGTTCCTGTGTTCCTTACCTTTC 83

Qy   936 TGTCCCTGTGTATAGATTATGTAAAGCCTTGTGTAAATATGAGATGTGTCAAAATGAT 995
         ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    84 TGTCCCTGTGTATAGATTATGTAAAGCCTTGTGTAAATATGAGATGTGTCAAAATGAT 143

Qy   996 GCAGTAAATGAGCAGTGACAATGACAGTGTACTGCGAGAGAAATTTACTCCTGCCTAGAACTGGAG 1055
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   144 GCAGTAAATGAGCAGTGACAATGACAGTGTACTGCGAGAG-AAATTTACTCCTGCCTAGAACTGGAG 202

Qy  1056 GGTTTTATGGGTCTGTAATTTCCCACACTCATTGCTGAAAGCTTAATTAAGTACTTCA 1115
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   203 GGTTTTTATGGGTCTGTAATTTCCCACACTCATTGCTGAAAGCTTAATTAAGTACTTCA 262

Qy  1116 AAAACGTATCTCCATTGTTTTACCTTCTTGAGGGGAACCGGTCTTGTTAACCAGCCCTGAG 1175
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   263 AAAACGTATCTCCATTGTTTTACCTTCTTGAGGGGAACCGGTCTTGTTAACCAGCCCTGAG 322

Qy  1176 TTGTCTACCCCAAACAATCCTGTCGTCATTTTCAAAGATGCAAAATGGTGTTATTATTGT 1235
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   323 TTGTCTACCCCAAACAATCCTGTCGTCATTTTCAAAGATGCAAAATGGAGTTATTGAATTGT 382

Qy  1236 CTCCACCATTGTCACACACAGGAATGCCTAATAATAGCAACCCTTGTCTCCCTCTCT 1295
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   383 CTCCACCATTGTCACACACAGGAATGCCTAATAATAGCAACCCTTGTCTCCCTCTCTCT 442

Qy  1296 CCTTTGCAAATGGCTCAGTGACTGCGAAGAGGCCGACTAATAGCCAGAGTTAAATATAAAT 1355
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   443 CCTTTGCAAATGGCTCAGTGACTGCGAAGAGGCGGACTAATAGCCAGAGTTAAATATAAAT 502

Qy  1356 ACAAATTAATAATAACATAGAGAACAGCAGCCAATACCAGAAAAAAAGAATTCTTGGTAAATGAT 1415
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   503 ACAAATTAATAATAACATAGAGAACAGCAGCCAATACCAGAAAAAAAGAATTCTGGTAAATGAT 562

Qy  1416 GTGAAAAATTGACAGCTCCCTCACTCTTAAGGTTGCTGCTATATACAGTCTAGGTTTTCT 1475
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   563 GTGAAAAATTGACAGCTCCCTCACTCTTAAGGTTGCTGCTATATACAGTCTAGGTTTTCT 622

Qy  1476 GTTTGGAAATAGTAGGTAGGGTAAAATCTAAGACCTGCACAAGGGCAGTGAGAGACATTTACA 1535
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   623 GTTTGGAAATAGGTAGGGTAAAATCTAAGACCTGCACAAGGGCAGTGAGAGACATTTACA 682

Qy  1536 G-CCTTCCTCTCTATTTGTTTTTTAAGGAAAAGTCAACTCC--TGAAATGTCCCCTTAGCT 1592
         | ||||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db   683 GCCCTCCTCTCTATTTGTTTTTTTTAAGGAAAGTCAACTCCCCTGAAATGTCCCCTTAGCT 742

Qy  1593 ATAATCAGAAAACTAAGAACATTATTCTGTGTCAACAATGTATTTATGGAGAAGTAAA 1652
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   743 ATAATCAGAAAACTAAGAACATTATTCTGTGTCCACAATGTATTTATGAAGAGAAAAATAAA 802

Qy  1653 AATAAGTTCCACAGCAACACAAAAACATGAATTAT 1687
         ||||||||||||||||||||||     ||||||||
Db   803 AATAAGTTCCACAGCAACACATTTACATGAATTAT 837
```

RESULT 12

Matches 719; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    49 GTGTTGGGAGACTTCAGGTGTTGGGAAATCTTCGTTAGTCCATCCTATGCCAAAATCAA 108
         | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GGGTTGGGAGACTTCAGGTGTTGGGAAATCTTCGTTAGTCCATCCTATGCCAAAATCAA  60

Qy   109 GTGCTGGGAAATCATCATGGACTGTGGCGTCGTCAGTGGATGTCAGTTCATGATTAC 168
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 GTGCTGGGAAATCATCATGGACTGTGGCGTCGTCAGTGGATGTCAGTTCATGATTAC 120

Qy   169 AAAGAAGGAACCCAGAGAGAGAGACCTACTACAGAATTATGGGATGTTGGAGGCTCT 228
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AAAGAAGGAACCCAGAGAGAGAGACCTACTACAGAATTATGGGATGTTGGAGGCTCT 180

Qy   229 GTGGGCAGTGCCAGCAGCGTGAAAAGCACAAGAGCAGTATTCTCACAACTCCGTAAATGGT 288
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTGGGCAGTGCCAGCAGCGTGAAAAGCACAAGAGCAGTATTCTCACAACTCCGTAAATGGT 240

Qy   289 ATTATTTTCGTACACGACTTAACAAATAAGAAGTCCTCCCAAAACTTCGGCTCGTTGGTCA 348
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ATTATTTTCGTACACGACTTAACAAATAAGAAGTCCTCCCAAAACTTCGGCTCGTTGGTCA 300

Qy   349 TTGGAAGCTCTCAACAGGGATTGGTGCCAACTGGAGTCTTGGTGACAAATGGGGATTAT 408
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TTGGAAGCTCTCAACAGGGATTGGTGCCAACTGGAGTCTTGGTGACAAATGGGGATTAT 360

Qy   409 GATCAAGAACAGTTGCTGATAACCAAATACCACTGTGTAATAGGGACTAAACTGGAC 468
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 GATCAAGAACAGTTGCTGATAACCAAATACCACTGTGTAATAGGGACTAAACTGGAC 420

Qy   469 CAGATTCATGAAACAAAGCGCCATGAAGTTTAACTAGGACTGCTTCCTGGCTGAGGAT 528
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CAGATTCATGAAACAAAGCGCCATGAAGTTTAACTAGGACTGCTTCCTGGCTGAGGAT 480

Qy   529 TTCAATCCAGAGAAATTAATTTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCT 588
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TTCAATCCAGAGAAATTAATTTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCT 540

Qy   589 TCCAATGCTGTCAAGCTCAGTAGTTTTTGATAAGGTCATAGAAGAAGAGATACTTTTA 648
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TCCAATGCTGTCAAGCTCAGTAGTTTTTGATAAGGTCATAGAAGAAGAGATACTTTTA 600

Qy   649 AGAGAAGTAATCAGATTCCAGGCTTTCCTGATCGGAAAAGATTTGGGCAGGAACATTA 708
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 AGAGAAGTAATCAGATTCCAGGCTTTCCTGATCGGAAAAGATTTGGGCAGGAACATTA 660

Qy   709 AAGAGCCTTCATTATGACTGAATTACACTCATCCTTTGGAAGAGTGAGCAAGCAGTGGCA 768
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 AAGAGCCTTCATTATGACTGAATTACACTCATCCTTTGGAAGAGTGAGCAAGCAGTGGCA 720
```

RESULT 14
BX099226
LOCUS   BX099226   718 bp   mRNA   linear   EST 06-FEB-2003
DEFINITION BX099226 Soares_fetal liver spleen 1NFLS_S1 Homo sapiens cDNA clone
            IMAGp998I061023 ; IMAGE:435149, mRNA sequence.
ACCESSION   BX099226
VERSION   BX099226.1 GI:27843984
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 718)
AUTHORS   Ebert,L., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M.,
            Radelof,U., Schneider,D. and Korn,B.
TITLE   Human UnigeneSet - RZPD3
JOURNAL   Unpublished (2003)
COMMENT   Contact: Ina Rolfs
            RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
            Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
            RZPD; IMAGp998I061023.
            RZPDLIB; I.M.A.G.E. cDNA Clone Collection;

```
Db   421 ||||||||||||||||||||||||||||||||||||||||||||||||||||||
         GTTGGTAATAGGGACTAAACTGGATCAGATTCAGATCATGAAGAACCAAGCCGCGCCATGAAGTTTTAAC 480

Qy   504 TAGGACTGCTTCCTGGCTGAGGATTCAATCCAGAGAAATTAATTTGGACTGCACAAA 563
         ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db   481 TAGGACTGCTTCCTGGCTGAGGATTCAATCCAGAGAAATTAATTTGGATTGCACAAA 540

Qy   564 TCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCAGTAGTTTTTGATAA 623
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCAGTAGTTTTTGATAA 600

Qy   624 GGTCATAGAGAAGAGATACTTTTTAAGAGAAGTAATCAGATTCCAGGCTTTCCTGATCG 683
         ||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db   601 GGTCATAGAGAAGAGATACTTTTTAAGAGAAGATGGTAATCAGATTCCAGGCTTTCCTGATCG 660

Qy   684 GAAAAGATTTGGGGAGGAACATTAAAGAGGCCTTCATTATGACTAAAGAGCCTTCATTATGACTCATCCT 743
         |||||||||||||||||||||||||||||||||   |||||||||| ||||||||||||||||
Db   661 AAAAAGGTTTGGGGAGGAACATTAAAGAGGCCTTACATTATGACTGAATTACACTCATCCT 720

Qy   744 TTGGAAGGTGAGCAAGCAGTGGCAGTTTTCACAGC-TCATCTTGCTGTGTTGTCAATTAT 802
         |||||||  |||  ||||||||||||||||||||| | |||||||||||  |||||||||
Db   721 TTGGAAGCTCTCAACAGGGATTGGTGCCAACTGGAGTCTTGGTGACAAATGGGGATTAT 780

Qy   803 TACCATCACAGCCTTTAACAAAATCATCTTAAAATGCTACCCTTCCAGCCTTTACCCTTTA 862
         |||||||||||||| |||||||||||||| ||| |||||||||||||||||||||||||||
Db   781 TCCTGTCCCAGCCTTTTAGCAAAATCATCTCTAAAATGCTACCCTTCCAGCCCTGCCGTTTA 840

Qy   863 ATGGAAAAATGAAAG 877
         |||||||| ||||||
Db   841 ATGGAAAACTGAAAG 855
```

RESULT 13
CN361073
LOCUS   CN361073   720 bp   mRNA   linear   EST 16-MAY-2004
DEFINITION 328775509 GRN_ES Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN361073
VERSION   CN361073.1 GI:47361007
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 720)
AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
            Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
TITLE   Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT   Contact: Brandenberger R
            Geron Corporation
            Regenerative Medicine
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 720   Std Error: 0.00.
FEATURES   Location/Qualifiers
     source   1..720
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cells, cell lines H1, H7, and
                     H9"
                     /clone_lib="GRN_ES"
                     /note="oligo dT primed, full-length enriched cDNA library
                     from undifferentiated hES cell lines H1 (p32), H7 (p29),
                     and H9 (p26) maintained in feeder-free conditions"
ORIGIN

Query Match   42.4%; Score 718.4; DB 7; Length 720;
Best Local Similarity   99.9%; Pred. No. 3.4e-177;

```
Human UnigeneSet - RZPD3 (RZPDLIB No.972)
http://www.rzpd.de/CloneCards/cgi-
bin/showLib.pl.cgi/response?libNo=972 Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13r, Primer sequence: TTTCACAGGAAACAGCTATGAC.
```

FEATURES
    source   1..718
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="IMAGp998I061023 ; IMAGE:435149"
               /sex="male"
               /dev_stage="20 week-post conception fetus"
               /lab_host="DH10B (ampicillin resistant)"
               /clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
               /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen 1NFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5'
AACTGGAAGAATTAATTAAGATCTTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN

Query Match   42.4%;   Score 718;   DB 5;   Length 718;
Best Local Similarity   100.0%;   Pred. No. 4.3e-177;
Matches 718;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

```
Qy    46 CTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAAT 105
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAAT 60

Qy   106 CAAGTGCTGGGAAATCCATCATGAACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGAT 165
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CAAGTGCTGGGAAATCCATCATGAACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGAT 120

Qy   166 TACAAAGAAGGAACCCCAGAAGAGAAGAGACCTACTACATAGAATTATGGGATGTTGGAGGC 225
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TACAAAGAAGGAACCCCAGAAGAGAAGAGACCTACTACATAGAATTATGGGATGTTGGAGGC 180

Qy   226 TCTGTGGGCAGTGCCAGCAGCGCGTGAAAAGCAGAGAGAAGCAGTATTCTACAACTGGTAAAT 285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TCTGTGGGCAGTGCCAGCAGCGCGTGAAAAGCAGAGAGAAGCAGTATTCTACAACTCCGTAAAT 240

Qy   286 GGTATTATTTTCGTACACGACTTAACAAATAAGAGAGTCCTCCCAAAACTTGCGTCGTTGG 345
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GGTATTATTTTCGTACACGACTTAACAAATAAGAGAGTCCTCCCAAAACTTGCGTCGTTGG 300

Qy   346 TCAATTGGAAGCTCTCAACAGGGATTTGGTGCCAACTGGAGTCTTGGTGACAAATGGGGAT 405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 TCAATTGGAAGCTCTCAACAGGGATTTGGTGCCAACTGGAGTCTTGGTGACAAATGGGGAT 360

Qy   406 TATGATCAAGAACAGTTGCTGATAACCACCACTGTTGGTAATAGGGACTAAACTG 465
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TATGATCAAGAACAGTTGCTGATAACCACCACTGTTGGTAATAGGGACTAAACTG 420

Qy   466 GACCAGATTCATGAAACAAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCTGCTGAG 525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GACCAGATTCATGAAACAAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCTGCTGAG 480

Qy   526 GATTTCAATCCAGAAGAAATTAATTGGACTGCACAAATCCACCGTACTTAGTTGCAGGT 585
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GATTTCAATCCAGAAGAAATTAATTGGACTGCACAAATCCACCGTACTTAGTTGCAGGT 540
```

RESULT 15
BM451532

LOCUS   BM451532   989 bp   mRNA   linear   EST 05-FEB-2002
DEFINITION   AGENCOURT_6395013 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5493017
           5', mRNA sequence.
ACCESSION   BM451532
VERSION   BM451532.1   GI:18500572
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 989)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: ATCC
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Agencourt Bioscience Corporation
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           http://image.llnl.gov
           Plate: LLAM12116 row: a   column: 18
           High quality sequence stop: 652.
FEATURES   Location/Qualifiers
    source   1..989
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="IMAGE:5493017"
               /tissue_type="retinoblastoma"
               /lab_host="DH10B (phage-resistant)"
               /clone_lib="NIH_MGC_67"
               /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."

ORIGIN

Query Match   42.1%;   Score 712.4;   DB 4;   Length 989;
Best Local Similarity   96.6%;   Pred. No. 1.4e-175;
Matches 749;   Conservative   0;   Mismatches 22;   Indels 4;   Gaps   2;

```
Qy    63 AGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTCTGGGAAATCC 122
          |||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db   182 AGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGTGGGAAATCC 241

Qy   123 ATCATGGACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAGGAACCCC 182
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   242 ATCATGGACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAGGAACCCC 301

Qy   183 AGAAGAGAAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAG 242
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   302 AGAAGAGAAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAG 361

Qy   243 CAGCTGAAAGCACAAGAGCACAGAGCAGTATTCTACAACTCCGTAAATGTATTATTTCGTACA 302
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   243 CAGCTGAAAGCACAAGAGCACAGAGCAGTATTCTACAACTCCGTAAATGTATTATTTCGTACA 302
```

```
Qy   586 TCTTCCAATCGTGTCAAGCTCAGTAGGTTTTTGATAAGGTCATAGAGAAGAGATACTTT 645
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TCTTCCAATCGTGTCAAGCTCAGTAGGTTTTTGATAAGGTCATAGAGAAGAGATACTTT 600

Qy   646 TTAAGAGAAGTAATCAGATTCCAGGCTTCCTGATCGGAAAAGATTTGGGCGCAGGAACA 705
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TTAAGAGAAGTAATCAGATTCCAGGCTTCCTGATCGGAAAAGATTTGGGCGCAGGAACA 660

Qy   706 TTAAAGAGACCCTCATTATGACTGAATTACACTCATCCTTTGGAAGAGTGAGCAAGCAG 763
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 TTAAAGAGACCCTTCATTATGACTGAATTACACTCATCCTTTGGAAGAGTGAGCAAGCAG 718
```

```
Db         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       362 CAGCGTGAAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACA 421

Qy     303 CGACTTAACAAATAAGAAGTCCTCCCAAAACTTGGGTCGTTGGTCATTGGAAGCTCTCAA 362
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 CGACTTAACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAA 481

Qy     363 CAGGGATTTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTT 422
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 CAGGGATTTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTT 541

Qy     423 TGCTGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAAC 482
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     542 TGCTGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAAC 601

Qy     483 AAAGCGCCATGAAGTTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGA 542
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     602 AAAGCGCCATGAAGTTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGA 661

Qy     543 AATTAATTTGGACTGCACAAATCCACGGTACTTAGCTGCGCCAATGCTGTCAA 602
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     662 AATTAATTTGGACTGCACAAATCCACGGTACTTAGCTGCGCCAATGCTGTCAA 721

Qy     603 GCTCAGTAGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAATCA 662
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     722 GCTCAGTAGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAATCA 781

Qy     663 GATTCCAGGC-TTTCCTGATCGGAAAAGATTGGGGCAGGAACATTAAAGAGCCTTCATT 721
           ||||||||||   |||||||||||||||||||||||||||||||||||||||
Db     782 GATTCCAGGCTTTTCCTGATCGGGGAAAGATTGGGGCAGGAACATTAAAGAGCCTTCATT 841

Qy     722 ATGACTGAATTACACTCATCCTTTGGAAGAGTGAGCAGTGGCAGTTTTTCACAGCT 781
           ||||||||||||||||||||||||||||||||||||||||  |||||||||||||
Db     842 ATGACTGAATTACACTCATCCTTTGGAAGAGTGAGCAGTGCCAGTTTTTCACAGCT 901

Qy     782 CATCTTG---CTGTGTTCAATTATTACCATCACAGCCTTTTAACAAAATCATCTT 833
           ||||| ||   |||||||||||    | |||| |   |||  |||||||||||
Db     902 CATCCTGGCTGGGGNTCAATTATTTATTTACCATCACCGCCTTTTAACCAAAATCTT 956
```

Search completed: June 30, 2005, 16:53:29
Job time : 6332.94 secs

THIS PAGE BLANK (USPTO)

```
                    GenCore version 5.1.6
             Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:     June 30, 2005, 14:24:03 ; Search time 7762.11 Seconds
                                        (without alignments)
                                       10574.845 Million cell updates/sec

Title:            US-09-945-173-1
Perfect score:    1694
Sequence:         1 agggaaggcagtggcaagat.........aaacatgaattattgaacta 1694

Scoring table:    IDENTITY NUC
                  Gapop 10.0 , Gapext 1.0
Searched:         4708233 seqs, 2422760795 residues

Total number of hits satisfying chosen parameters:        9416466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      GenEmbl:*
             1: gb_ba:*
             2: gb_htg:*
             3: gb_in:*
             4: gb_om:*
             5: gb_ov:*
             6: gb_pat:*
             7: gb_ph:*
             8: gb_pl:*
             9: gb_pr:*
            10: gb_ro:*
            11: gb_sts:*
            12: gb_sy:*
            13: gb_un:*
            14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

Result No.	Score	% Query Match	Query Length	DB	ID	Description
1	1694	100.0	1694	6	AX465669	AX465669 Sequence
2	1680.6	99.2	3362	9	HSM807705	BX647559 Homo sapi
3	1680.6	99.2	3449	9	HSM807739	BX647593 Homo sapi
4	1666.8	98.4	1999	9	BC020832	BC020832 Homo sapi
5	1660.8	98.0	3352	9	HSM807810	BX647664 Homo sapi
6	1387	81.9	2550	9	AK025772	AK025772 Homo sapi
7	1022.2	60.3	65838	9	AC133474	AC133474 Homo sapi
8	1022.2	60.3	146010	9	AC009289	AC009289 Homo sapi
c 9	875.6	51.7	2064	10	AC050194	BC050194 Mus muscu
10	711	42.0	711	6	AX465671	AX465671 Sequence
11	488.2	28.8	3327	5	AJ720020	AJ720020 Gallus ga
12	425	25.1	1064	5	BC073035	BC073035 Xenopus l
13	397	23.4	1040	5	BC078191	BC078191 Danio rer
14	395.4	23.3	1291	5	AY391440	AY391440 Danio rer
15	374.8	22.1	385	2	BD076273	BD076273 5' EST of
c 16	330.8	19.5	155418	2	AC117754	AC117754 Mus muscu
17	321	18.9	321	6	AX906158	AX906158 Sequence
18	321	18.9	321	6	BD041691	BD041691 Sequence
19	301.8	17.8	245326	2	AC107288	AC107288 Rattus no

20	301.8	17.8	266597	2	AC119722	AC119722 Rattus no
21	223	13.2	223	6	CQ719549	CQ719549 Sequence
22	139.2	8.2	66009	2	AC101250	AC101250 Mus muscu
23	135	8.0	119730	9	AC117472	AC117472 Homo sapi
24	135	8.0	188485	2	AC027454	AC027454 Homo sapi
c 25	102.2	6.0	165836	5	AC138153	AC138153 Danio rer
26	102.2	6.0	174601	5	BX571724	BX571724 Zebrafish
c 27	92.6	5.5	66009	2	AC139727	AC139727 Danio rer
c 28	81.6	4.8	66009	2	AC101250	AC101250 Mus muscu
c 29	77.6	4.6	87017	2	AC140043	AC140043 Mus muscu
c 30	77.6	4.6	29949	2	AC112488	AC112488 Homo sapi
c 31	67.2	4.0	179351	2	AC143587	AC143587 Macaca mu
c 32	66.8	3.9	188485	2	AC027454	AC027454 Homo sapi
33	66.8	3.9	919	6	CQ581943	CQ581943 Drosophil
34	60	3.5	934	6	AY070945	AY070945 Drosophil
35	56.8	3.4	1087	6	AR508948	AR508948 Sequence
36	56.8	3.4	60	6	CQ540888	CQ540888 Sequence
37	56.8	3.4	3048	6	CQ581942	CQ581942 Sequence
38	56.8	3.4	3215	6	CQ591803	CQ591803 Sequence
39	56.8	3.4	3888	6	CQ591806	CQ591806 Sequence
40	56.8	3.4	8610	3	AC012956	AC012956 Drosophil
41	56.8	3.4	177576	3	AC022349	AC022349 Drosophil
c 42	56.8	3.4	179077	3	AC018489	AC018489 Drosophil
43	56.8	3.4	300469	3	AE003503	AE003503 Drosophil
44	51.8	3.1	1141	6	AX083744	AX083744 Sequence
c 45	51	3.0	146917	2	CR450801	CR450801 Danio rer

```
                              ALIGNMENTS

RESULT 1
AX465669
LOCUS       AX465669               1694 bp    DNA        linear   PAT 16-JUL-2002
DEFINITION  Sequence 1 from Patent WO0218425.
ACCESSION   AX465669
VERSION     AX465669.1  GI:21899929
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Meyers,R.A.
  TITLE     Patent: WO 021842S-A 1 07-MAR-2002;
  JOURNAL   Millennium Pharmaceuticals (US)
            47324, a human g-protein and uses therefor
FEATURES             Location/Qualifiers
     source          1..1694
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS             19..729
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD42379.1"
                     /db_xref="GI:21899930"
                     /translation="MASLDRVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVD
                     VRVHDYKEGTPERKTYYIELWDVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKS
                     SQNLRRWSLEALNRDLVPTGVLVTNGDYDDEQFADNQIPLLVIGTKLDQIHETKRHEV
                     LTRTAFLAEDFNPEEINLDCTNPRYLAAGSSNAVKLSRFFDKVIEKRYFLREGNQIPG
                     FPDRKRFGAGTLKSLHYD"

ORIGIN
         Query Match          100.0%;  Score 1694;  DB 6;  Length 1694;
         Best Local Similarity  100.0%;  Pred. No. 0;
         Matches 1694;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy        1 AGGGAAGGCAGTGGCAAGATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGAC 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 AGGGAAGGCAGTGGCAAGATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGAC 60

Qy       61 TCAGGTGTTGGGAAATCTTCGTTAGTCCATCTTCCTATGCCAAATCAAGTGCTGGGAAAT 120
```

```
Db    61 TCAGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTAGTCCAAATCAAGTGCTGGGAAAT  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   121 CCATCATGGACTGTGGCTGCTCAGTGGATGTCAGAGTTCAGAGTTCATGATTACAAAGAGGAACC  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CCATCATGGACTGTGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAGGAACC  180

Qy   181 CCAGAAGAGAAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCGTGGGCAGTGCC  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CCAGAAGAGAAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCGTGGGCAGTGCC  240

Qy   241 AGCAGCGGTGAAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTA  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AGCAGCGGTGAAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTA  300

Qy   301 CACGACTTAACAAATAAGAGAGTCCTCCCAAAACTTGCGTCGTCGTTGGTCATTGGAAGCTCTC  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CACGACTTAACAAATAAGAGAGTCCTCCCAAAACTTGCGTCGTCGTTGGTCATTGGAAGCTCTC  360

Qy   361 AACAGGGATTGGTGCCAACTGGAGTCTTCGGTAATAGGGATTATGGGACCAGATTCATGAA  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AACAGGGATTGGTGCCAACTGGAGTCTTCGGTAATAGGGATTATGGGACCAGATTCATGAA  420

Qy   421 TTTGCTGATAACCAAATACCACTGTTGGTAACTAGGACTGCTTTCCTGGCTGAGATTTCAATCCAGAA  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 TTTGCTGATAACCAAATACCACTGTTGGTAACTAGGACTGCTTTCCTGGCTGAGATTTCAATCCAGA  480

Qy   481 ACAAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGATTTCAATCCAGAA  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ACAAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGATTTCAATCCAGA  540

Qy   541 GAAATTAATTGGACTGCACAAATCCACGTACTTAGCTGCAGGTTCTTCCCAATGCTGTC  600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GAAATTAATTGGACTGCACAAATCCACGTACTTAGCTGCAGGTTCTTCCCAATGCTGTC  600

Qy   601 AAGCTTCAGTAGGTTTTTTGATAAGGTCATAGAGAGAAGAGATACTTTTTAAGAGAAGTAAT  660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 AAGCTTCAGTAGGTTTTTTGATAAGGTCATAGAGAGAAGAGATACTTTTTAAGAGAAGTAAT  660

Qy   661 CAGATTCCAGCTTTCCTGATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGCCTTCAT  720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CAGATTCCAGCTTTCCTGATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGCCTTCAT  720

Qy   721 TATGACTGAATTACACTCATCCTTTGGAAGAGTGAGCAAGCCAGTGCAGTTTTTCACAGC  780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 TATGACTGAATTACACTCATCCTTTGGAAGAGTGAGCAAGCCAGTGCAGTTTTTCACAGC  780

Qy   781 TCATCTTGCTGTGTCAATTATTACCATCACAGCCTTTAACAAAATCATCTAAAATGC  840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 TCATCTTGCTGTGTCAATTATTACCATCACAGCCTTTAACAAAATCATCTAAAATGC  840

Qy   841 TACCCTTCAGCCTTACCCTTACCCTTTAATGAAAAATGAAAGGAGTGACAATACGGGGAGTCCA  900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 TACCCTTCAGCCTTACCCTTACCCTTTAATGAAAAATGAAAGGAGTGACAATACGGGGAGTCCA  900

Qy   901 AACTTTGTCCCTGTTCTCTGTGTTCCTTACCTTTCTGTCCCTGTGTATAGATTATGTAAA  960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 AACTTTGTCCCTGTTCTCTGTGTTCCTTACCTTTCTGTCCCTGTGTATAGATTATGTAAA  960

Qy   961 AGCCTTGTGTAAATATGAGATGTTGTCAAAATGATGCAGTAAAATGACAGTGTA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 AGCCTTGTGTAAATATGAGATGTTGTCAAAATGATGCAGTAAAATGACAGTGTA 1020

Qy  1021 CTGCAGAGAAAATTTACTCTTGCCTAGAACTGGAGGGTTTTTATGGGTCTGTAATTTTCC 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 CTGCAGAGAAAATTTACTCTTGCCTAGAACTGGAGGGTTTTTATGGGTCTGTAATTTTCC 1080

Qy  1081 CACACTCATTGCTGAAAGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTACCT 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 CACACTCATTGCTGAAAGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTACCT 1140

Qy  1141 TCTTGAGGGGAACGGTCTTGTTGTTAACCAGCCCTGAGTTGTCTCACCACCATTGTCACCACCAGGAAT 1200
```

```
Db  1141 TCTTGAGGGGAACGGTCTTGTTGTTAACCAGCCCTGAGTTGTCTCACCACCATTGTCACCACCAGGAAT 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1201 ATTTTCAAAGATGCAAAAATGGTGTTATTTAAGTGCTCCACCATTGTCACCACCAGGAAT 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 ATTTTCAAAGATGCAAAAATGGTGTTATTTAAGTTGTCCCACCATTGTCACCACCAGGAAT 1260

Qy  1261 GCCTAATAATGCAACCCTTGTCTCCTCCTTTGCAAATGGCTCAGTGACTGG 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 GCCTAATAATGCAACCCTTGTCTCCTCCTTTGCAAATGGCTCAGTGACTGG 1320

Qy  1321 AAGAGGCGGACTAATAGCCAGAGTTAAATTAATAATATACATAGAGACA 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 AAGAGGCGGACTAATAGCCAGAGTTAAATTAAATACAAATTAATACATAGAGACA 1380

Qy  1381 GCAATACCAGAAAAAAAGAATTCTGTAAAATGATGTGAAAAATTGACAGCTCCCTCACT 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 GCAATACCAGAAAAAAAGAATTCTGTAAAATGATGTGAAAAATTGACAGCTCCCTCACT 1440

Qy  1441 CTTAAGGTTGCTGCTATATACAGTCTAGGTTTCTGTTGGAAATAGGTAGGTAAAATC 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 CTTAAGGTTGCTGCTATATACAGTCTAGGTTTCTGTTGGAAATAGGTAGGTAAAATC 1500

Qy  1501 TAAGACCTGCACAAGGGCAGTGAGAGACATTCACAGCCTCCTCTCCATTTGTTTTTTAA 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 TAAGACCTGCACAAGGGCAGTGAGAGACATTCACAGCCTCCTCTCTATTTGTTTTTTTAA 1560

Qy  1561 GGAAAAGTCAACTCCTGAAATGTCCCTTAGCTATAATCAGAAAACTAAGAATTATTATTCT 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 GGAAAAGTCAACTCCTGAAATGTCCCTTAGCTATAATCAGAAAACTAAGAATTATTATTCT 1620

Qy  1621 GTGTCAACAATGTATTTATTGAGAGAAGTAAAAATAAGTTCCACAGCAACACAAAAACAT 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621 GTGTCAACAATGTATTTATTGAGAGAAGTAAAAATAAGTTCCACAGCAACACAAAAACAT 1680

Qy  1681 GAATTATTGAACTA 1694
            ||||||||||||||
Db  1681 GAATTATTGAACTA 1694
```

```
RESULT 2
HSM807705
LOCUS       HSM807705              3362 bp    mRNA    linear   PRI 30-AUG-2003
DEFINITION  Homo sapiens mRNA; cDNA DKFZp313A1611 (from clone DKFZp313A1611).
ACCESSION   BX647559
VERSION     BX647559.1  GI:34366716
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3362)
  AUTHORS   Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  CONSRTM   The German Human cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by BMFZ (Biomedical Research Center at the
            Heinrich-Heine-University, Duesseldorf/Germany) within the cDNA
            sequencing consortium of the German Genome Project. This clone
            (DKFZp313A1611) is available at the RZPD in Berlin. Please contact
            the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES             Location/Qualifiers
     source          1..3362
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp313A1611"
```

```
             /tissue_type="human bone marrow"
             /clone_lib="313 (synonym: hlcc2)). Vector pTriplEx2; host
             DH10B; sites SfiIA + SfiIB"
             /dev_stage="adult"
polyA_signal  3308..3313
polyA_site    3332..3313
ORIGIN

Query Match      99.2%;   Score 1680.6;  DB 9;   Length 3362;
Best Local Similarity  99.8%;  Pred. No. 0;
Matches 1683;   Conservative   0;  Mismatches  4;  Indels  0;  Gaps  0;

Qy     1 AGGGAAGGCAGTGGCAAGATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGAC    60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    10 AGGGAAGGCAGTGGCAAGATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGAC    69

Qy    61 TCAGGTGTTGGGAAATCTTCGTTAGTCCATCGTCCTATGCCAAAATCAAGTCGTGGGAAAT   120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    70 TCAGGTGTTGGGAAATCTTCGTTAGTCCATCGTCCTATGCCAAAATCAAGTCGTGGGAAAT   129

Qy   121 CCATCATGGACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATGATTACAAAGAGGAACC   180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   130 CCATCATGGACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATGATTACAAAGAGGAACC   189

Qy   181 CCAGAGAGAAGAACCTACTACATAGAATTATGGGATGTTGGAGGCTCGTGTGGGCAGTGCC   240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   190 CCAGAGAGAAGAACCTACTACATAGAATTATGGGATGTTGGAGGCTCGTGTGGGCAGTGCC   249

Qy   241 AGCAGCGTGAAAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTA   300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   250 AGCAGCGTGAAAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTA   309

Qy   301 CACGACTTAACAAATAGAGAAGTCCTCCCAAAACTTGGCTCGTTGGTCATTGGAAGCTCTC   360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   310 CACGACTTAACAAATAGAGAAGTCCTCCCAAAACTTGGCTCGTTGGTCATTGGAAGCTCTC   369

Qy   361 AACAGGATTTGGTGCCAACTGGAGTCTTGTGAGTGGGATTATGATCAAGAACAG   420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   370 AACAGGATTTGGTGCCAACTGGAGTCTTGTGAGTGGGATTATGATCAAGAACAG   429

Qy   421 TTTGCTGATAACCAAATACCACTGTTGGTAATAGGACTAAACTGGACCAGATTCATGAA   480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   430 TTTGCTGATAACCAAATACCACTGTTGGTAATAGGACTAAACTGGACCAGATTCATGAA   489

Qy   481 ACAAAGCGCCATGACTGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAA   540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   490 ACAAAGCGCCATGACTGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAA   549

Qy   541 GAAATTAATTTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCTTCCCAATGCTGTC   600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   550 GAAATTAATTTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCTTCCCAATGCTGTC   609

Qy   601 AAGCTCAGTAGGTTTTTTTGATAAGGTCATAGAGAGAGATACTTTTTTAAGAAGAAGGTAAT   660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   610 AAGCTCAGTAGGTTTTTTTGATAAGGTCATAGAGAGAGATACTTTTTTAAGAAGAAGGTAAT   669

Qy   661 CAGATTCCAGGCTTCCAGCTTTCCTGATCGGAAAAGAGATTTGGGCAGGAACATTAAAGAGCCTTCAT   720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   670 CAGATTCCAGGCTTCCAGCTTTCCTGATCGGAAAAGAGATTTGGGCAGGAACATTAAAGAGCCTTCAT   729

Qy   721 TATGACTGAATTACACTCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTTCACAGC   780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   730 TATGACTGAATTACACTCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTTCACAGC   789

Qy   781 TCATCTTGCTGTGTTCAATTATTACCATCACAGCCTTTTAACAAAATCATCTTAAAATGC   840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   790 TCATCTTGCTGTGTTCAATTATTACCATCACAGCCTTTTAACAAAATCATCTTAAAATGC   849

Qy   841 TACCCTTCAGCCTTACCCTTTAATGAAAAATGAAAGGAAGTGACAATACCGGGAGGTCCA   900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   850 TACCCTTCAGCCTTACCCTTTAATGAAAAATGAAAGGAAGTGACAATACCGGGAGGTCCA   909

Qy   901 AACTTTGTCCCTGTCCCTGTGTCCTTACCTTCGTGTCCCTGTATAGAATTATGTGTAAA   960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   910 AACTTTGTCCCTGTCCCTGTGTCCTTACCTTCGTGTCCCTGTATAGAATTATGTGTAAA   969

Qy   961 AGGCCTTGTGTAAATATGAGATGTTGTCAAAATGACAGTGAAATGAGCAATGACAGTGTA   1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   970 AGGCCTTGTGTAAATATGAGATGTTGTCAAAATGACAGTGAAATGAGCAATGACAGTGTA   1029

Qy  1021 CTGCAGAGAAAATTACTTCTTGCCTAGAACTGAAGGGTTTTATGGGTCTGTAATTTCC   1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1030 CTGCAGAGAAAATTACTTCTTGCCTAGAACTGAAGGGTTTTATGGGTCTGTAATTTTCC   1089

Qy  1081 CACACTCATTGCTGAAAGCTTAATTAAGTACTTCAAAAACGTATCCCATTGTTTACCT   1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1090 CACACTCATTGCTGAAAGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTACCT   1149

Qy  1141 TCTTGAGGGGACGGTCTTGTTAACCAGCCCTGAGTTGTCTACCCCAAACAATCTCTGTC   1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1150 TCTTGAGGGGACGGTCTTGTTAACCAGCCCTGAGTTGTCTACCCCAAACAATCTCTGTC   1209

Qy  1201 ATTTTCAAAGATGCAAAATGGTGTTATTAATTGTCTCCACCATTGTCACACACAGGAAT   1260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1210 ATTTTCAAAGATGCAAAATGGTGTTATTAATTGTCTCCACCATTGTCACACACAGGAAT   1269

Qy  1261 GCCTAATAATAGCAACCCCTTGTCTCCCCTCCCCTTTGCAAATGGCTCAGTGACTCGG   1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1270 GCCTAATAATAGCAACCCCTTGTCTCCCCTCCCCTTTGCAAATGGCTCAGTGACTCGG   1329

Qy  1321 AAGAGGCCGACTAATAGCCAGAGTTAAATAATAATTAATAATACATAGAGAACA   1380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1330 AAGAGGCCGACTAATAGCCAGAGTTAAATAATAACAAATTAATAATACATAGAGAACA   1389

Qy  1381 GCAATACCAGAAAAAAAGAATTCTGGTAAAATGATGTGAAAAATGAGCAGCTCCCTCACT   1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1390 GCAATACCAGAAAAAAAGAATTCTGGTAAAATGATGTGAAAAATGAGCAGCTCCCTCACT   1449

Qy  1441 CTTAAGGTTGCTGCTATATACAGTCTAGGTTTTCTGTTGGAAATAGGTAGGGTAAAATC   1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1450 CTTAAGGTTGCTGCTATATACAGTCTAGGTTTTCTGTTGGAAATAGGTAGGGTAAAATC   1509

Qy  1501 TAAGACCTGCACAAGGGCAGCATTACAGCCTCCTCCTATTTGTTTTTTAA   1560
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1510 TAAGACCTGCACAAGGGCAGCAGTGAGAGACATTACAGCCTCCTCCTATTTGTTTTTTTAA   1569

Qy  1561 GGAAAAGTCAACTCCCTGAAAATGTCCCTAGCTATAATCAGAAAACTAAGAATATTATTCT   1620
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1570 GGAAAAGTCAACTCCCTGAAAATGTCCCTAGCTATAATCAGAAAACTAAGAATATTATTCT   1629

Qy  1621 GTGTCAACAATGTATTTATGGAGAAGTAAAAATAAGTTCCACAGCAACACAAAAACAT   1680
         ||||||||||||||||||||||||||||||||||||||||||||||||||  ||||
Db  1630 GTGTCAACAATGTATTTATGGAGAAGTAAAAATAAGTTCCACAGCAACACATTACAT   1689

Qy  1681 GAATTAT   1687
         |||||||
Db  1690 GAATTAT   1696
```

RESULT 3
HSM807739
LOCUS   HSM807739   3449 bp   mRNA   linear   PRI 30-AUG-2003
DEFINITION   Homo sapiens mRNA; cDNA DKFZp313M1333 (from clone DKFZp313M1333).
ACCESSION   BX647593
VERSION   BX647593.1 GI:34366750
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3449)
  AUTHORS   Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
             Fobo,G., Han,M. and Wiemann,S.
  CONSRTM   The German Human cDNA Consortium
  TITLE   Direct Submission
  JOURNAL   Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764

COMMENT   Neuherberg, GERMANY
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by BMFZ (Biomedical Research Center at the
Heinrich-Heine-University, Duesseldorf/Germany) within the cDNA
sequencing consortium of the German Genome Project. This clone
(DKFZp313M133) is available at the RZPD in Berlin. Please contact
the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.

FEATURES   Location/Qualifiers
source   1..3449
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="DKFZp313M133"
            /tissue_type="human heart"
            /clone_lib="313 (synonym: hlcc2). Vector pTriplEx2; host
            DH10B; sites SfiIA + SfiIB"
            /dev_stage="adult"
polyA_signal   3311..3316
polyA_site   3336
ORIGIN

Query Match   99.2%;   Score 1680.6;   DB 9;   Length 3449;
Best Local Similarity 99.8%;   Pred. No. 0;
Matches 1683;   Conservative   0;   Mismatches   4;   Indels   0;   Gaps   0;

Qy   1 AGGGAAGGCAGTGGCAAGATGGCGTCCCTGGATCGGTGAAGTACTGGTGTTGGGAGAC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13 AGGGAAGGCAGTGGCAAGATGGCGTCCCTGGATCGGTGAAGTACTGGTGTTGGGAGAC 72

Qy   61 TCAGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAAT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73 TCAGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAAT 132

Qy   121 CCATCATGGACGTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGGAGGAACC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   133 CCATCATGGACGTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGGAGGAACC 192

Qy   181 CCAGAGAGAAGAGACTACTACTACATAGAATTATGGGATTGGAGGCTGTGGAGGCCAGTGCC 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   193 CCAGAGAGAAGAGACTACTACTACATAGAATTATGGGATTGGAGGCTGTGGAGGCCAGTGCC 252

Qy   241 AGCAGCGGTGAAAAGCACAAGACACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   253 AGCAGCGGTGAAAAGCACAAGACACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTA 312

Qy   301 CACGACTTAACAAATAAGAAGTCCTCCAACTGGAGTCGTCGTTGGTCATTGGAAGCTCTC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   313 CACGACTTAACAAATAAGAAGTCCTCCAACTGGAGTCGTCGTTGGTCATTGGAAGCTCTC 372

Qy   361 AACAGGGATTTGGTGCCAACTGGAGTCGTTGGTAATAGGGACTAAACTGACCAGATTCAAGAACAG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   373 AACAGGGATTTGGTGCCAACTGGAGTCGTTGGTAATAGGGACTAAACTGACCAGATTCAAGAACAG 432

Qy   421 TTTGCTGATACCAAATACCACTGTTGGTAATAGGGACTAAACTGACCAGATTCAATCCAGAA 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   433 TTTGCTGATACCAAATACCACTGTTGGTAATAGGGACTAAACTGACCAGATTCAATCCAGAA 492

Qy   481 ACAAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAA 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   493 ACAAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAA 552

Qy   541 GAAATTAATTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTC 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   553 GAATTTAATTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTC 612

Qy   601 AAGCTCAGTAGGTTTTTGATAAGGTCATAGGAGAGAGATACTTTTTAAGAAGAAGTAAT 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   613 AAGCTCAGTAGGTTTTTGATAAGGTCATAGGAGAGAGATACTTTTTAAGAAGAAGTAAT 672

Qy   661 CAGATTCCAGGCTTTCCTGATCGGAAAAGATTTGGGCAGGAACATTAAAGAGACCTTCAT 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   673 CAGATTCCAGGCTTTCCTGATCGGAAAAGATTTGGGCAGGAACATTAAAGAGACCTTCAT 732

Qy   721 TATGACTGAATTACACTCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTCACAGC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   733 TATGACTGAATTACACTCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTCACAGC 792

Qy   781 TCATCTGCTGTGTTCAATTATTACCATCACAGCCTTTAACAAAATCATCTAAAATGC 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   793 TCATCTGCTGTGTTCAATTATTACCATCACAGCCTTTAACAAAATCATCTAAAATGC 852

Qy   841 TACCCTTCAGCTTACCCTTTAATGAAAAATGAAAGGAAGTGACAATACGGGAGGTCCA 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   853 TACCCTTCAGCCTTACCCTTTAATGAAAAATGAAAGGAAGTGACAATACGGGAGGTCCA 912

Qy   901 AACTTTGTCCCTGTCTCTGTGTTCCTTTCTGTCCCTGTATAGATTATGTAAA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   913 AACTTTGTCCCTGTCTCTGTGTTCCTTTCTGTCCCTGTATAGATTATGTAAA 972

Qy   961 AGCCTTGTGTAAAATGAGATGTGTCAAAATGATGCAGTAAATGAGCAATGACAGTGTA 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   973 AGCCTTGTGTAAAATGAGATGTGTCAAAATGATGCAGTAAATGAGCAATGACAGTGTA 1032

Qy 1021 CTGCAGAGAAAATTTACTCTTGCCTAGAACTGGAGGGTTTTTTATGGGTCTGTAATTTTCC 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1033 CTGCAGAGAAAATTTACTCTTGCCTAGAACTGGAGGGTTTTTTATGGGTCTGTAATTTTCC 1092

Qy 1081 CACACTCATTGCTGAAAAGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTACCT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1093 CACACTCATTGCTGAAAAGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTACCT 1152

Qy 1141 TCTTGAGGGGAACGGTCTTGTTAACCAGCCCTGAGTTGTCTCACCCCAAACAATCTCTGTC 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1153 TCTTGAGGGGAACGGTCTTGTTAACCAGCCCTGAGTTGTCTCACCCCAAACAATCTCTGTC 1212

Qy 1201 ATTTTCAAAGATGCAAAATGTGTTATTTAATTGTCTCCACCATTGTCACACACAGGAAT 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1213 ATTTTCAAAGATGCAAAATGTGTTATTTAATTGTCTCCACCATTGTCACACACAGGAAT 1272

Qy 1261 GCCTAATAATAGCAACCCTTGTCTCCCCTCTCTCCCTTGCAAATGGCTCAGTGACTGG 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1273 GCCTAATAATAGCAACCCTTGTCTCCCCTCTCTCCCTTGCAAATGGGTCAGTGACTGG 1332

Qy 1321 AAGAGGCGGACTAATAGCCAGAGTTAAAATTAATAATACAAATTAATAATAATACAGAGAACA 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1333 AAGAGGCGGACTAATAGCCAGAGTTAAAATTAATAATACAAATTAATAATAATACAGAGAACA 1392

Qy 1381 GCAATACCAGAAAAAAGAGAATTCTGGTAAAATGATGTGAAAAATTGACAGCTCCCTCACT 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1393 GCAATACCAGAAAAAAGAGAATTCTGGTAAAATGATGTGAAAAATTGACAGCTCCCTCACT 1452

Qy 1441 CTTAAGGTTGCTGCTATATACAGTCTAGGTTTTCTGTTTGGAAATAGGTAGGGTAAAATC 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1453 CTTAAGGTTGCTGCTATATACAGTCTAGGTTTTCTGTTTGGAAATAGGTAGGGTAAAATC 1512

Qy 1501 TAAGACCTGCACAAGGGCAGTGAGAAGTGTCCCTTAGCTATAATCAGAAAACTAAGAATATTATTCT 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1513 TAAGACCTGCACAAGGGCAGTGAGAAGTGTCCCTTAGCTATAATCAGAAAACTAAGAATATTATTCT 1572

Qy 1561 GGAAAAGTCAACTCCTGAAAATGTCCCTGAAAATGTCCCTTAGCTATAATCAGAAAACATAAGAATATTATTCT 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1573 GGAAAAGTCAACTCCTGAAAATGTCCCTGAAAATGTCCCTTAGCTATAATCAGAAAACTAAGAATATTATTCT 1632

Qy 1621 GTGTCAACAATGTATTTATGGAGAGAAGTAAAAATAAGTTCCACAGCAACACAAAAACAT 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1633 GTGTCAACAATGTATTTATGGAGAGAAGTAAAAATAAGTTCCACAGCAACACATTTACAT 1692

Qy 1681 GAATTAT 1687
           |||||||
Db 1693 GAATTAT 1699

RESULT 4
BC020832
LOCUS   BC020832   1999 bp   mRNA   linear   PRI 29-JUN-2004
DEFINITION   Homo sapiens RAB, member of RAS oncogene family-like 3, mRNA (cDNA
             clone MGC:23920 IMAGE:4771781), complete cds.
ACCESSION   BC020832
VERSION   BC020832.1 GI:18088571
KEYWORDS   MGC.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1999)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
             Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
             Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
             Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
             Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
             Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
             Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
             Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
             Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
             McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
             Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
             Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
             Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
             Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
             Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
             Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
             Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
             Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
             human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2 (bases 1 to 1999)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (03-JAN-2002) National Institutes of Health, Mammalian
             Gene Collection (MGC), Cancer Genomics Office, National Cancer
             Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
             USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: CLONTECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Sequencing Group at the Stanford Human Genome
             Center, Stanford University School of Medicine, Stanford, CA 94305
             Web site: http://www-shgc.stanford.edu
             Contact: (Dickson, Mark) mcd@paxil.stanford.edu
             Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
             R. M.

             Clone distribution: MGC clone distribution information can be found
             through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
             Series: IRAL Plate: 37 Row: m Column: 23
             This clone was selected for full length sequencing because it
             passed the following selection criteria: matched mRNA gi: 28376659.
FEATURES   Location/Qualifiers
     source   1..1999
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:23920 IMAGE:4771781"
                     /tissue_type="Testis, embryonal carcinoma"
                     /clone_lib="NIH_MGC_61"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
     gene   1..1999
                     /gene="RABL3"
                     /note="synonym: MGC23920"

     CDS
                     /db_xref="LocusID:285282"
                     10..720
                     /gene="RABL3"
                     /codon_start=1
                     /product="RAB, member of RAS oncogene family-like 3"
                     /protein_id="AAH20832.1"
                     /db_xref="GI:18088572"
                     /db_xref="LocusID:285282"
                     /translation="MASLDRVKLVLVLGDSGVGKSSLVHLLCQNVLGNPSWTVGCSVD
                     VRVHDYKEGTPEEKTCYIELWDVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKS
                     SQNLRRWSLEALNRDLVPTGVLTNGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEV
                     LTTAFLAEDFNPEEINLDCTNPRYLAAGSSNAVKLSRFFDKVIEKRYFLRENQIPG
                     PPDRKRFGAGTLKSLHYD"

ORIGIN
Query Match   98.4%;   Score 1666.8;   DB 9;   Length 1999;
Best Local Similarity   99.6%;   Pred. No. 0;
Matches 1671; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   10 AGTGGCCAAGATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTT 69
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AGTGGCCAAGATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTT 60

Qy   70 GGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGG 129
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATG 120

Qy   130 ACTGTGGGCTGCTCAGTGGATGCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAG 189
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAG 180

Qy   190 AAGACCTACTACATAGAATTATGGGATGTGGAGGTCTGTGGGCAGTGCCAGCAGCGTG 249
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AAGACCTGCTACATAGAATTATGGGATGTGGAGGTCTGTGGGCAGTGCCAGCAGCGTG 240

Qy   250 AAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGACTTA 309
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGACTTA 300

Qy   310 ACAAATAAGAAGTCCTCCAAAACTTGCGTGCTTGGTCATTGGAAGCTCAACAGGGAT 369
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ACAAATAAGAAGTCCTCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCAACAGGGAT 360

Qy   370 TTGGTGCCAACTGGAGTCTTGGTGACAAATGGGGATTATGATGATCAAGAACAGTTGCTGAT 429
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TTGGTGCCAACTGGAGTCTTGGTGACAAATGGGGATTATGATGATCAAGAACAGTTTGCTGAT 420

Qy   430 AACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACTAAACTGGAAACAAAGCGC 489
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 AACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACTAAACTGGAAACAAAGCGC 480

Qy   490 CATGAAGTTTAAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAAT 549
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CATGAAGTTTAAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAAT 540

Qy   550 TTGGACTGCACAAATCCACGGTACTACTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCAGT 609
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TTGGACTGCACAAATCCACGGTACTACTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCAGT 600

Qy   610 AGTTTTTTGATAAGGTCATAGAGAGAAGAGATACTTTTTAAGAGAAGAGGTAATCAGATTCCA 669
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 AGTTTTTTGATAAGGTCATAGAGAGAAGAGATACTTTTTAAGAGAAGAGGTAATCAGATTCCA 660

Qy   670 GGCTTTCCTGATCGGAAAAGATTTGGGGCCAGGAACATTAAAGAGGCCTTCATTATGACTGA 729
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 GGCTTTCCTGATCGGAAAAGATTTGGGGCCAGGAACATTAAAGAGCCTTCATTATGACTGA 720

Qy   730 ATTACACTCATCCTTTGGAAGAGTGGAGCAAGCAGTGGCAGTTTTTCACAGCTCATCTTGC 789
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 ATTACACTCATCCTTTGGAAGAGTGGAGCAAGCAGTGGCAGTTTTTCACAGCTCATCTTGC 780

Qy   790 TGTGTTCAATTATTACCATCACAGCCTTTTAACAAAATCATCTTAAAATGCTACCCTTCA 849
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db    781  TGTGTTCAATTATTACCATCACAGCCTTTAACAAATCATCTTAAAATGCTACCCTTCA  840

Qy    850  GCCTTACCCTTAATGGAAAAATGAAAGGAAGTGACAATACGGGAGGTCCAAACTTTGTC  909
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  GCCTTACCCTTAATGGAAAAATGAAAGGAAGTGACAATACGGGAGGTCCAAACTTTGTC  900

Qy    910  CCTGTTCTCGTGTTCCTTACCTTTCGTCCCTGTGTATAGATTATGTAAAAGCCTTGTG  969
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  CCTGTTCTCGTGTTCCTTACCTTTCGTCCCTGTGTATAGATTATGTAAAAGCCTTGTG  960

Qy    970  TAAATATGAGATGTTGTCAAAATGATGCAGTAAATGAGCAATGACAGTGTACTGCGAGA  1029
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961  TAAATATGAGATGTTGTCAAAATGATGCAGTAAATGAGCAGTGTACTGCGAGA  1020

Qy   1030  AAATTTACTCTTGCCTAGAACTGGAGGGTTTTATGGGTCTGTAATTTTCCCACACTCAT  1089
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021  AAATTTACTCTTGCCTAGAACTGGAGGGTTTTATGGGTCTGTAATTTTCCCACACTCAT  1080

Qy   1090  TGCTGAAAGCTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTACCTTCTTGAGGG  1149
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081  TGCTGAAAGCTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTACCTTCTTGAGGG  1140

Qy   1150  GAACGGTCTTGTGTTAACCAGCCCTGAGTTGTCTACCCCAAACAATCTCTGTCATTTCAAA  1209
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141  GAACGGTCTTGTGTTAACCAGCACTGAGTTGTCTACCCCAAACAATCTCTGTCATTTTCAAA  1200

Qy   1210  GATGCAAAATGGTGTTAATTGTCTCCACCATTGTCACACAGGAATGCCTAATAA  1269
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201  GATGCAAAATGGTGTTATTTAATTGTCTCCACCATTGTCACACAGGAATGCCTAATAA  1260

Qy   1270  TAGCAACCCCTGTCTCCCTCCTTTGCAAATGGGTCAGTGACTGGAAGAGGCGG  1329
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261  TAGCAACCCCTGTCTCCCTCCTTTGCAAATGGGTCAGTGACTGGAAGAGGCGG  1320

Qy   1330  ACTAATAGCCAGAGTTAAATATACAAATTAATAATACATAGAGAACACAGCAATACCA  1389
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321  ACTAATAGCCAGAGTTAAATATACAAATTAATAATACATAGAGAACACAGCAATACCA  1380

Qy   1390  GAAAAAAAGAATTCTGTAAAATGATGTGAAAAATTGACAGCTCCCTCACTCTTAAGGTT  1449
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381  GAAAAAAAGAATTCTGTAAAATGATGTGAAAAATTGACAGCTCCCTCACTCTTAAGGTT  1440

Qy   1450  GCTGCTATATACAGTCTAGGTTTTCTGTTTGGAAATAGGTAGGGTAAAATCTAAGACCTG  1509
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441  GCTGCTATATACAGTCTAGGTTTTCTGTTTGGAAATAGGTAGGGTAAAATCTAAGACCTG  1500

Qy   1510  CACAAGGGCAGTGAGAGACATTTACAGCCTCCTATTTGTTTTTTGTTTTTTAAGGAAAAGTC  1569
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501  CACAAGGGCAGTGAGAGACATTTACAGCCTCCTATTTGTTTTTTTGTTTTTTAAGGAAAAGTC  1560

Qy   1570  AACTCCTGAAAATGTCCCTTAGCTATAAATCAGAAAAACTAAGAATATTATTCTGTGTCAACA  1629
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561  AACTCCTGAAAATGTCCCTTAGCTATAAATCAGAAAAACTAAGAATATTATTCTGTGTCAACA  1620

Qy   1630  ATGTATTTATGGAGAGAAGTAAAAATAAGTTCCAGCAACACAAAAACATGAATTAT  1687
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621  ATGTATTTATGGAGAGAAGTAAAAATAAGTTCCAGCAACACAACATTTACATGAATTAT  1678
```

```
FEATURES             Location/Qualifiers
     source          1..3352
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp313O171"
                     /tissue_type="human fetal skin"
                     /clone_lib="313 (synonym: hlcc2). Vector pTriplEx; host
                     DH10B; sites SfiIA + SfiIB"
                     /dev_stage="fetal"
     polyA_signal    3296..3301
     polyA_site      3321
ORIGIN

Query Match             98.0%;  Score 1660.8;  DB 9;  Length 3352;
Best Local Similarity   99.5%;  Pred. No. 0;
Matches 1676;  Conservative  0;  Mismatches  7;  Indels  1;  Gaps  1;

Qy      5  AAGGCAGTGGCAAGATGGCGTCCCTGATCGGTGGTACTGGTGTTGGGAGACTTCAG  64
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  AAGGCAGTGGCAAGATGGCGTCCCTGATCGGTGGTACTGGTGTTGGGAGACTTCAG  60

Qy     65  GTGTTGGGAAATCTTCGTTAGTCCATCTCCATGCCCAAATCAAGTGCTGGGAAATCCAT  124
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  GTGTTGGGAAATCTTCGTTAGTCCATCTCCATGCCCAAAATCAAGTGCTGGGAAATCCAT  120

Qy    125  CATGGACTGTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAG  184
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  CATGGACTGTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAG  180

Qy    185  AAGAGAAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGCCAGTGCCAGCA  244
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  AAGAGAAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCCAGTGCCAGCA  240

Qy    245  GGCGTGAAAAAGCAACAAGAGCAGTAGTTCTACAACTCCGTAAAATGTATTATTTCGTACACG  304
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  GGCGTGAAAAAGCAACAAGAGCAGTAGTTCTACAACTCCGTAAAATGTATTATTTCGTACACG  300

Qy    305  ACTTAACAAATAAGAAGTACTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACA  364
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  ACTTAACAAATAAGAAGTACTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACA  360

Qy    365  GGGATTTGGTGCCAACTGGAGTCTTCGTGACAAAATGGGATTATGATGATCAAGAACAGTTTG  424
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  GGGATTTGGTGCCAACTGGAGTCTTCGTGACAAAATGGGATTATGATGATCAAGAACAGTTTG  420

Qy    425  CTGATAACCAAATACCACCTGTTGGTAATAGGACTAAACTGGACCAGATTCATGAAACAA  484
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  CTGATAACCAAATACCACCTGTTGGTAATAGGACTAAACTGGACCAGATTCATGAAACAA  480

Qy    485  AGCGCCATGAAGTTTTAACTAGGACTGCTTTCCTGCTGAGGATTTCAATCCAGAAGAAA  544
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  AGCGCCATGAAGTTTTAACTAGGACTGCTTTCCTGCTGAGGATTTCAATCCAGAAGAAA  540

Qy    545  TTAATTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGC  604
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  TTAATTTGGACTGCACAAATCCACGGTACTTAGCAGCAGGTTCTTCCAATGCTGTCAAGC  600
```

```
RESULT 5
HSM807810
LOCUS       HSM807810               3352 bp    mRNA            linear   PRI 30-AUG-2003
DEFINITION  Homo sapiens mRNA; cDNA DKFZp313O171 (from clone DKFZp313O171).
ACCESSION   BX647664
VERSION     BX647664.1  GI:34366821
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3352)
  AUTHORS   Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
```

```
                                                              Db   1681 TTAT 1684
                                                                        ||||
                                                              1681 TTAT 1684

RESULT 6
LOCUS       AK025772                2550 bp    mRNA    linear   PRI 13-SEP-2003
DEFINITION  Homo sapiens cDNA: FLJ22119 fis, clone HEP18852.
ACCESSION   AK025772
VERSION     AK025772.1  GI:10438390
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
            Okamoto,S., Okitani,R., Ota,T., Suzuki,M., Obayashi,M., Nishi,T.,
            Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2550)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Laboratory of Genome Structure Analysis, Human
            Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
            Japan (E-mail:flcdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing; Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Departent of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES             Location/Qualifiers
     source          1..2550
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HEP18852"
                     /cell_line="HepG2"
                     /cell_type="hepatoma"
                     /clone_lib="HEP"
                     /note="cloning vector pME18SFL3"
ORIGIN

Query Match      81.9%;  Score 1387;  DB 9;  Length 2550;
Best Local Similarity   99.6%;  Pred. No. 0;
Matches 1390;  Conservative   0;  Mismatches   5;  Indels   0;  Gaps   0;
```

```
Qy    293 TTTTCGTACGACTTAACAAATAAGAGAGTCCTCCAAAACTTGCGTCGTTGGTCATTGG 352
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TTTTCGTACGACTTAACAAATAAGAGAGTCCTCCAAAACTTGCGTCGTTGGTCATTGG 60

Qy    353 AAGCTCTCAACAGGGATTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGATC 412
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AAGCTCTCAACAGGGATTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGATC 120

Qy    413 AAGAACAGTTGCTGATAACCAAATACCACTGTTGGTAATAGGACTAAACTGGACCAGA 472
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AAGAACAGTTGCTGATAACCAAATACCACTGTTGGTAATAGGACTAAACTGGACCAGA 180

Qy    473 TTCATGAACAAAGCGCCATGAAGTTTAACTAGGATTTTAACTAGGATTTATTCTGTG 532
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TTCATGAACAAAGCGCCATGAAGTTTAACTAGGATTTTAACTAGGATTTATTCTGTG 240

Qy    533 ATCCAGAAGAAATTAATTTGGACTGCACAAATCCACGTACTTAGCTGCAGGTTCTTCCA 592
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    241 ATCCAGAAGAAATTAATTTGGACTGCACAAATCCGCGGTACTTAGCTGCAGGTTCTTCCA 300
```

```
Qy    605 TCAGTAGGTTTTGATAAGGTCATAGAGAGAGAGATACTTTTAAGAGAAGGTAATCAGA 664
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TCAGTAGGTTTTGATAAGGTCATAGAGAGAGAGATACTTTTAAGAGAAGGTAATCAGA 660

Qy    665 TTCCAGGCTTCCTGATCGGAAAAGAGATTGGGGCAGGAACATTAAAGAGCCTTCATTATG 724
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TTCCAGGCTTCCTGATCGGAAAAGAGATTGGGGCAGGAACATTAAAGAGCCTTCATTATG 720

Qy    725 ACTGAATTACACTCATCCTTGGAAGAGTGAGCAAGCAGTGCAGTTTTCACAGCTCAT 784
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 ACTGAATTACACTCATCCTTGGAAGAGTGAGCAAGCAGTGCAGTTTTCACAGCTCAT 780

Qy    785 CTTGCTGTGTTCAATTATTACCATCACAGCCTTTAACAAAATCATCTTAAAATGCTACC 844
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CTTGCTGTGTTCAATTATTACCATCACAGCCTTTAACAAAATCATCTTAAAATGCTACC 840

Qy    845 CTTCAGGCCTTTACCCTTTAATGCAATGAAAGAGGAGTGACAATACGGGAGGTCCAAACT 904
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CTTCAGCCCTTACCCTTTAATGCAATGAAAGAGGAGTGACAATACGGGAGGTCCAAACT 900

Qy    905 TTGTCCCGTGTCTCTGTGTTCCTTACCTTTCTGTCCCGTGTATAGATTATGTAAAAGCC 964
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TTGTCCCGTGTCTCTGTGTTCCTTACCTTTCTGTCCCGTGTATAGATTATGTAAAAGCC 960

Qy    965 TTGTGTAAATATGAGATGTGTCAGTAAATGATGAGCAGCAATGACAGTGTACTGC 1024
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TTGTGTAAATATGAGATGTGTCAGTAAATGATGAGCAGCAATGACAGTGTACTGC 1020

Qy   1025 AGAGAAAATTACTCTTGCCTAGAACTGGAGGGTTTTATGGGTCTGTAATTTTCCCACA 1084
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AGAGAAAATTACTCTTGCCTAGAACTGGAGGGTTTTATGGGTCTGTAATTTTCCCACA 1080

Qy   1085 CTCATTGCTGAAAGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTACCTTCTT 1144
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CTCATTGCTGAAAGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTATCTTCTT 1140

Qy   1145 GAGGGGAACGGTCTTCTTGTAACCAGCCCGAGTGTCTCCACCCCAAACAATCTCTGTCATTT 1204
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 GAGGGGAACGGTCTTCTTGTAACCAGCCCGAGTGTCTCCACCCCAAACAATCTCTGTCATTT 1200

Qy   1205 TCAAAGATGCAAAATGGTGTTATTAATTGTCTCCACCATGTCACACACA-GGAATGCC 1263
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   1201 TCAAAGATGCAAAATGGTGTTATTAATTGTCTCCACCATGTCACACAGGGAATGCC 1260

Qy   1264 TAATAATAGCAACCCTGTCTCCCCTGTCTCCCTGTTAAATATACAAATTAATATACATAGAGAACACAGCA 1323
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 TAATAATAGCAACCCTGTCTCCCCTGTCTCCCTGTTAAATATACAAATTAATATACATAGAGAACACAGCA 1320

Qy   1324 AGGCCGGACTAATAGCCAGAGTTAAATATAAATAATCAAATTAATAATACATAGAGAACAGCA 1383
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 AGGCCGGACTAATAGCCAGAGTTAAATATAAATAATCAAATTAATAATACATAGAGAACAGCA 1380

Qy   1384 ATACCAGAAAAAAAGAATTCTGGTAAATGTAAAATGATGTGAAAAATTGACAGCTCCCTCACTCTT 1443
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 ATACCAGAAAAAAAGAATTCTGGTAAATGTAAAATGATGTGAAAAATTGACAGCTCCCTCACTCTT 1440

Qy   1444 AAGGTTGCTGCTATATACAGTCTAGGTT 1503
          ||||||||||||||||||||||||||||
Db   1441 AAGGTTGCTGCTATATACAGTCTAGGTT 1500

Qy   1504 GACCTGCACAAGGGCAGTGAGAGACATTACAGCCTTCCTCTCTATTTGTTTTTTAAGGA 1563
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 GACCTGCACAAGGGCAGTGAGAGACATTACAGCCTTCCTCTCTATTTGTTTTTTAAGGA 1560

Qy   1564 AAAGTCAACTCCTGAAAATGTCCCTTAGCTATAATCAGAAAACTAAGAATATATTCTGTG 1623
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 AAAGTCAACTCCTGAAAATGTCCCTTAGCTATAATCAGAAAACTAAGAATATATTCTGTG 1620

Qy   1624 TCAACAATGATTTATGGAGAGAAGTAAAAATAAGTTCCACGCAACACCAACAAAAACATGAA 1683
          ||||||||||||||||||||||||||||||||||||||||||  |||||||||||||||||
Db   1621 TCAACAATGATTTATGGAGAGAAGTAAAAATAAGTTCCACAGCCACCAACAACATTACATGAA 1680

Qy   1684 TTAT 1687
```

```
Qy          1381 ATTTACATGAATTAT 1395
                 |||||||||||||
Db
```

```
RESULT 7
AC133474
LOCUS       AC133474               65838 bp    DNA     linear   PRI 03-DEC-2002
DEFINITION  Homo sapiens 3 BAC RP11-719N16 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC133474
VERSION     AC133474.2  GI:23334664
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 65838)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Albrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Barnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Louisegeg,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma.J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P. Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newton,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,L.,
            Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,D., Wang,Q., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
            Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 65838)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 65838)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 65838)
```

```
Qy   593 ATGCTGTCAAGCTCAGTAGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTAAGAG 652
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ATGCTGTCAAGCTCAGTAGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTAAGAG 360

Qy   653 AAGGTAATCAGATTCCAGGCTTCCTGATCGGAAAAGATTGGGCAGGAACATTAAAGA 712
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AAGGTAATCAGATTCCAGGCTTCCTGATCGGAAAAGATTGGGCAGGAACATTAAAGA 420

Qy   713 GCCTTCATTATGACTGAATTACACTCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAGTTT 772
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GCCTTCATTATGACTGAATTACACTCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAGTTT 480

Qy   773 TTCACAGCTCATCTTGCTGTGTTCAATTATTACCATCACAGCCTTTAACAAAATCATCT 832
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TTCACAGCTCATCTTGCTGTGTTCAATTATTACCATCACAGCCTTTAACAAAATCATCT 540

Qy   833 TAAAATGCTACCCTTCAGCCTTAATGAAAAATGAAAGGAAGTGACAATACGG 892
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TAAAATGCTACCCTTCAGCCTTAATGAAAAATGAAAGGAAGTGACAATACGG 600

Qy   893 GAGGTCCAAACTTTGTCCCGTGTTCTCTGTGTTCCTTACCTTTCTGTCCCGTGTATAGAT 952
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   601 GAGGTCCAAACTTTGTCCCGTGTTCTCTGTGTTCCTTACCTTTCTGTCCCGTGTGTAGAT 660

Qy   953 TATGTAAAAGCCTTGTGTAAATATGAGATGTTGTCAAAATGAGTAAATGAGCAATG 1012
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 TATGTAAAAGCCTTGTGTAAATATGAGATGTTGTCAAAATGAGTAAATGAGCAATG 720

Qy  1013 ACAGTGTACTGCAGAGAAAATTACTCTTGCCTAGAACTGGAGGGTTTTTATGGGTCTGT 1072
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 ACAGTGTACTGCAGAGAAAATTACTCTTGCCTAGAACTGGAGGGTTTTTATGGGTCTGT 780

Qy  1073 AATTTTCCCACACTCATTGCTGAAAGCTTAATTAAGTAGTACTTCAAAAAACGTATCTCCATTG 1132
         ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   781 AATTTTCCCACACTCATTGCTGAAAGCTTAATTAAGTAGTACTTCAAAAAACGTATCTCCATTG 840

Qy  1133 TTTTACCTTCTTGAGGGAACGGTCGTTGTTAACCAGCCCTGAGTTGTCTACCCCAAACAA 1192
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 TTTTACCTTCTTGAGGGGAACGGTCGTTGTTAACCAGCCCTGAGTTGTCTACCCCAAACAA 900

Qy  1193 TCTCTGTCATTTTCAAAGATGCAAAATGGTGTTATTAATTAATTGTCTCCACCATTGTCACAC 1252
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 TCTCTGTCATTTTCAAAGATGCAAAATGGTGTTATTAATTAATTGTCTCCACCATTGTCACAC 960

Qy  1253 ACAGGAATGCCTAATAATAGCAACCCTTGTCTCCCCTTTGCAAATGCTCA 1312
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 ACAGGAATGCCTAATAATAGCAACCCTTGTCTCCCCTTTGCAAATGCTCA 1020

Qy  1313 GTGACTGGAAGAGGCGGACTAATAGCCAGAGTAAATATAACAAATTAATAATACAT 1372
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 GTGACTGGAAGAGGCGGACTAATAGCCAGAGTAAATATAACAAATTAATAATACAT 1080

Qy  1373 AGAGAACAGCAATACCAGAAAAAAAAGAATTCTGGTAAAATTGACAGCT 1432
         |||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 AGAGAACAGCAATACCAGAAAAAAAAGAATTCTGGTAAAATTGACAGCT 1140

Qy  1433 CCCTCACTCTTAAGGTTGCTGCTATATACAGTCTAGGTTTTCTGTTTGGAAATAGGTAGG 1492
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 CCCTCACTCTTAAGGTTGCTGCTATATACAGTCTAGGTTTTCTGTTTGGAAATAGGTAGG 1200

Qy  1493 GTAAAATCTAAGACCTGACAGGGCAGTGAGAGACATTTACAGCCTCCTCTATTTGT 1552
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 GTAAAATCTAAGACCTGACAGGGCAGTGAGAGACATTTACAGCCTCCTCTATTTGT 1260

Qy  1553 TTTTTTAAGGAAAAGTCAACTCCTGAAAATGTCCCTTAGCTATAATCAGAAAAACTAAGAAT 1612
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 TTTTTTAAGGAAAAGTCAACTCCTGAAAATGTCCCTTAGCTATAATCAGAAAAACTAAGAAT 1320

Qy  1613 ATTATTCTGTGTCAACAATGTATTTATGGAGAAGTAAAAAAATAAGTTCCAGCAACAC 1672
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 ATTATTCTGTGTCAACAATGTATTTATGGAGAAGTAAAAAAATAAGTTCCAGCAACAC 1380

Qy  1673 AAAAACATGAATTAT 1687
```

AUTHORS　　Worley,K.C.
TITLE　　　Direct Submission
JOURNAL　　Submitted (03-DEC-2002) Human Genome Sequencing Center, Department
　　　　　　of Molecular and Human Genetics, Baylor College of Medicine, One
　　　　　　Baylor Plaza, Houston, TX 77030, USA
COMMENT　　On Sep 27, 2002 this sequence version replaced gi:22830213.
　　　　　　INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
　　　　　　gc-help@bcm.tmc.edu

　　　　　　CLONE LENGTH: This sequence does not necessarily represent the
　　　　　　entire insert of this clone. Overlapping regions of clones are only
　　　　　　sequenced and submitted once, so the sequence for the remainder of
　　　　　　the insert may be found in the record for the adjacent clones.
　　　　　　Overlapping clones are noted at the beginning and end of the
　　　　　　Features listing.

　　　　　　ANNOTATION OF FEATURES:
　　　　　　　STSs are identified using ePCR (Genome Res. 7:541-550) searches
　　　　　　of a local database that includes entries from dbSTS, GDB, and
　　　　　　local mapping efforts.
　　　　　　　Repeats are identified using RepeatMasker (A. Smit and P. Green,
　　　　　　unpublished.) for Human and Mouse sequences.
　　　　　　　Genes and Region of sequence similarity are identified by BLAST
　　　　　　(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
　　　　　　EST and cDNA sequences. Genes demonstrate at least two exons
　　　　　　flanked by consensus splice sites that maintained sequence
　　　　　　continuity across the splice junctions. Sequences that are not
　　　　　　identical matches are annotated as similar.

　　　　　　SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
　　　　　　standard of double strand coverage with a minimum of 2 clones and 2
　　　　　　reads with no ambiguities or 2 chemistries with a minimum of 2
　　　　　　clones and 3 reads with no ambiguities. If the sequence quality for
　　　　　　a region does not meet this standard, it will be indicated in the
　　　　　　annotation as Low Coverage.

　　　　　　QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
　　　　　　standards - estimated error rate less than 1 per 10,000 bases.
　　　　　　Reports of lowest quality individual bases and measures of base
　　　　　　quality are listed below. Description of the metrics can be found
　　　　　　at URL:
　　　　　　http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht
　　　　　　ml.

FEATURES　　　　　Location/Qualifiers
　　source　　　　1..65838
　　　　　　　　　/organism="Homo sapiens"
　　　　　　　　　/mol_type="genomic DNA"
　　　　　　　　　/db_xref="taxon:9606"
　　　　　　　　　/chromosome="3"
　　　　　　　　　/clone="RP11-719N16"
　　misc_feature　1..2005
　　　　　　　　　/note="overlaps bases 117724. .119726 of clone AC117472"
　　　　　　　　　/function="clone overlap"
　　repeat_region　complement(1. .420)
　　　　　　　　　/rpt_family="L2"
　　STS　　　　　　2133. .2353
　　　　　　　　　/standard_name="SHGC-68332"
　　repeat_region　2818. .3009
　　　　　　　　　/rpt_family="MIR"
　　repeat_region　complement(4474. .5354)
　　　　　　　　　/rpt_family="LIM4"
　　repeat_region　5358. .5382
　　　　　　　　　/rpt_family="(TTTTG)n"
　　repeat_region　complement(5383. .5642)
　　　　　　　　　/rpt_family="AluSx"
　　repeat_region　complement(5643. .7192)
　　　　　　　　　/rpt_family="LIM4"
　　repeat_region　7550. .7851
　　　　　　　　　/rpt_family="AluY"
　　repeat_region　7892. .8199
　　　　　　　　　/rpt_family="MIR"
　　repeat_region　complement(8480. .8586)
　　　　　　　　　/rpt_family="LIME"

　　repeat_region　9046. .9175
　　　　　　　　　/rpt_family="LIME"
　　repeat_region　complement(9498. .10116)
　　　　　　　　　/rpt_family="LIMC5"
　　repeat_region　11023. .11046
　　　　　　　　　/rpt_family="AT_rich"
　　repeat_region　complement(12233. .12522)
　　　　　　　　　/rpt_family="AluSg"
　　repeat_region　complement(12953. .13127)
　　　　　　　　　/rpt_family="MIR"
　　repeat_region　complement(13247. .13436)
　　　　　　　　　/rpt_family="MER5A"
　　repeat_region　13772. .13813
　　　　　　　　　/rpt_family="(TG)n"
　　repeat_region　complement(14893. .15027)
　　　　　　　　　/rpt_family="MIR"
　　repeat_region　16272. .16299
　　　　　　　　　/rpt_family="AT_rich"
　　repeat_region　17561. .17601
　　　　　　　　　/rpt_family="AT_rich"
　　repeat_region　17909. .18207
　　　　　　　　　/rpt_family="AluY"
　　repeat_region　complement(18719. .19081)
　　　　　　　　　/rpt_family="MER47A"
　　STS　　　　　　20760. .20949
　　　　　　　　　/standard_name="SHGC-77253"
　　repeat_region　21591. .21610
　　　　　　　　　/rpt_family="(TG)n"
　　repeat_region　22278. .22315
　　　　　　　　　/rpt_family="AT_rich"
　　repeat_region　22316. .22596
　　　　　　　　　/rpt_family="AluSx"
　　repeat_region　22597. .22634
　　　　　　　　　/rpt_family="(TAAA)n"
　　repeat_region　22957. .23107
　　　　　　　　　/rpt_family="AluJb"
　　repeat_region　23327. .23425
　　　　　　　　　/rpt_family="LTR29B"
　　repeat_region　23401. .23779
　　　　　　　　　/rpt_family="MER39"
　　repeat_region　complement(24441. .24627)
　　　　　　　　　/rpt_family="MER5A"
　　repeat_region　complement(25067. .25188)
　　　　　　　　　/rpt_family="HAL1"
　　repeat_region　25240. .25367
　　　　　　　　　/rpt_family="L2"
　　repeat_region　25508. .25664
　　　　　　　　　/rpt_family="AluJb"
　　repeat_region　25670. .25958
　　　　　　　　　/rpt_family="AluJo"
　　repeat_region　25959. .25985
　　　　　　　　　/rpt_family="AT_rich"
　　repeat_region　complement(26771. .26936)
　　　　　　　　　/rpt_family="MLT1H"
　　repeat_region　complement(27115. .27217)
　　　　　　　　　/rpt_family="MLT1H"
　　repeat_region　complement(28823. .29001)
　　　　　　　　　/rpt_family="MIR"
　　repeat_region　complement(29469. .29599)
　　　　　　　　　/rpt_family="L2"

Query Match　　　　60.3%;　Score 1022.2;　DB 9;　Length 65838;
Best Local Similarity　99.7%;　Pred. No. 2.1e-240;
Matches 1024;　Conservative　0;　Mismatches　3;　Indels　0;　Gaps　0;

Qy　　　661 CAGATTCCAGGCTTTCCTGATCGGAAAAGATTTGGGCAGGAACATTAAAGAGCCTTCAT 720
　　　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　20515 CAGATTCCAGGCTTTCCTGATCGGAAAAGATTTGGGCAGGAACATTAAAGAGCCTTCAT 20574

Qy　　　721 TATGACTGAATTACACTCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTTCACAGC 780
　　　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　20575 TATGACTGAATTACACTCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTTCACAGC 20634

```
Qy    781 TCATCTGCTGTGTCAATTATTACCATCACAGCCTTTAACAAAATCATCTTAAAATGC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20635 TCATCTGCTGTGTCAATTATTACCATCACAGCCTTTAACAAAATCATCTTAAAATGC 20694

Qy    841 TACCCTTCAGCTTACCCTTTAATGGAAAAATGAAAGGAGTGACAATACGGGAGTCCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20695 TACCCTTCAGCTTACCCTTTAATGGAAAAATGAAAGGAGTGACAATACGGGAGTCCA 20754

Qy    901 AACTTGTCCCGTTCTCTGTGTTCCTTACCTTTCTGTCCCTGTGTATAGATTATGTAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20755 AACTTTGTCCCGTTCTCTGTGTTCCTTACCTTTCTGTCCCTGTGTATAGATTATGTAAA 20814

Qy    961 AGCCTTGTGTAAAATGAGATGTTGTCAAAATGATGCAGTAAATGACAGTGTA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20815 AGCCTTGTGTAAAATGAGATGTTGTCAAAATGATGCAGTAAATGACAGTGTA 20874

Qy   1021 CTGCAGAGAAAATTTACTCTTGCCTAGAACTGGAGGGTTTTATGGGTCTGTAATTTTCC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20875 CTGCAGAGAAAATTTACTCTTGCCTAGAACTGGAGGGTTTTATGGGTCTGTAATTTTCC 20934

Qy   1081 CACACTCATTGCTGAAAGCTTAATTAAGTACTTCAAAAACGTATCTTCCATTGTTTTACCT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20935 CACACTCATTGCTGAAAGCTTAATTAAGTACTTCAAAAACGTATCTTCCATTGTTTTACCT 20994

Qy   1141 TCTTGAGGGGAACGGTCTTGTTAACCAGCCCTGAGTTGTCTCACCCCAAACAATCTCTGTC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  20995 TCTTGAGGGGAACGGTCTTGTTAACCAGCCCTGAGTTGTCTCACCCCAAACAATCTCTGTC 21054

Qy   1201 ATTTTCAAAGATGCAAAATGGTGTTATTTAATTGTCTTCCACCATTGTCACACACAGGAAT 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21055 ATTTTCAAAGATGCAAAATGGTGTTATTTAATTGTCTTCCACCATTGTCACACACAGGAAT 21114

Qy   1261 GCCTAATAATAGCAACCCCTTGTCTCCCCTCTTCTCTCCTTTGCAAATGGCTCAGTGACTGG 1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21115 GCCTAATAATAGCAACCCCTTGTCTCCCCTCTTCTCTCCTTTGCAAATGGCTCAGTGACTGG 21174

Qy   1321 AAGAGGGGACTAAATAGCCAGAGTTAAAATAAATACAAATTAATAATACATAGAGAACA 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21175 AAGAGGGCGGACTAAATAGCCAGAGTTAAAATAAATACAAATTAATAATACATAGAGAACA 21234

Qy   1381 GCAATACCAGAAAAAGAATTCTGGTAAAATGATGTGAAAAATTGACAGCTCCCTCACT 1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21235 GCAATACCAGAAAAAGAATTCTGTGTAAAATGATGTGAAAAATTGACAGCTCCCTCACT 21294

Qy   1441 CTTAAGGTTGCTGCTATATACAGTCTAGGTTTCTGTTGAAAATAGGTAGGGTAAAATC 1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21295 CTTAAGGTTGCTGCTATATACAGTCTAGGTTTTCTGTTGGAAATAGGTAGGGTAAAATC 21354

Qy   1501 TAAGACCTGCACAAGGGCAGTGAGAGACATTTACAGCCTCCTCTCCTCCTCATTTGTTTTTTTAA 1560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21355 TAAGACCTGCACAAGGGCAGTGAGAGACATTTACAGCCTCCTCTCCTTCTATTTTGTTTTTTTAA 21414

Qy   1561 GGAAAAGTCAACTCCTGAAATGTCCCTTAGCTATAATCAGAAAACTAAGAATATTATTTCT 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21415 GGAAAAGTCAACTCCTGAAATGTCCCTTAGCTATAATCAGAAAACTAAGAATATTATTTCT 21474

Qy   1621 GTGTCAACAATGTATTTATGGAGAGAGTAAAAATAAGTTCCACAGCAACACAAAAACAT 1680
          |||||||||||||||||||||||||||          ||||||||||||||||||||||
Db  21475 GTGTCAACAATGTATTTATGGAGAGAGTAAAAATAAGTTCCACAGCAACCAACACATTTACAT 21534

Qy   1681 GAATTAT 1687
          |||||||
Db  21535 GAATTAT 21541
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 146010)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-44J3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 146010)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
            Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
            Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
            Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
            Karatas,A., Lehoczky,J., Lieu,C., Locke,K., Macdonald,P.,
            Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
            Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
            Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
            Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
            Tesfaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
            Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 146010)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,P., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 24, 2000 this sequence version replaced gi:8389538.
            All repeats were identified using RepeatMasker.
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------------ Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ------------------ Project Information
            Center project name: L1073
            Center clone name: 44_J_3
            ------------------
FEATURES             Location/Qualifiers
     source          1..146010
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-44J3"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   118..238
                     /rpt_family="L1ME"
```

repeat_region   complement(561..1188)
                /rpt_family="L1MC5"
repeat_region   2086..2109
                /rpt_family="AT_rich"
repeat_region   complement(3296..3585)
                /rpt_family="AluSg"
repeat_region   complement(4016..4190)
                /rpt_family="MIR"
repeat_region   complement(4310..4499)
                /rpt_family="MER5A"
repeat_region   4835..4876
                /rpt_family="(TG)n"
repeat_region   complement(5956..6090)
                /rpt_family="MIR"
repeat_region   7328..7381
                /rpt_family="AT_rich"
repeat_region   8624..8678
                /rpt_family="AT_rich"
repeat_region   8972..9270
                /rpt_family="AluY"
repeat_region   complement(9782..10144)
                /rpt_family="MER47A"
repeat_region   12654..12673
                /rpt_family="(TG)n"
repeat_region   13315..13378
                /rpt_family="AT_rich"
repeat_region   13379..13659
                /rpt_family="AluSx"
repeat_region   13660..13697
                /rpt_family="(TAAA)n"
repeat_region   14020..14170
                /rpt_family="FRAM"
repeat_region   14390..14488
                /rpt_family="MER39"
repeat_region   14464..14842
                /rpt_family="LTR29B"
repeat_region   complement(15504..15690)
                /rpt_family="MER5A"
repeat_region   complement(16098..16251)
                /rpt_family="HAL1"
repeat_region   16303..16430
                /rpt_family="L2"
repeat_region   16571..16727
                /rpt_family="FRAM"
repeat_region   16733..17021
                /rpt_family="AluJo"
repeat_region   17022..17048
                /rpt_family="AT_rich"
repeat_region   complement(17834..17999)
                /rpt_family="MLT1H"
repeat_region   complement(18178..18280)
                /rpt_family="MLT1H"
repeat_region   complement(19886..20064)
                /rpt_family="MIR"
repeat_region   complement(20532..20662)
                /rpt_family="L2"
repeat_region   complement(20710..20822)
                /rpt_family="L2"
repeat_region   21289..21309
                /rpt_family="L2"
repeat_region   21318..21381
                /rpt_family="AT_rich"
repeat_region   complement(21377..21489)
                /rpt_family="MIR"
repeat_region   complement(21515..22019)
                /rpt_family="MER5A"
repeat_region   complement(23539..24609)
                /rpt_family="MLT1D"
repeat_region   complement(24828..24935)
                /rpt_family="MER11C"
repeat_region   27466..27711
                /rpt_family="L2"
repeat_region   28480..28514
                /rpt_family="MIR"

repeat_region   complement(28668..28796)
                /rpt_family="(CCTA)n"
repeat_region   29313..29982
                /rpt_family="MIR"
repeat_region   29986..30244
                /rpt_family="L1PA9"
repeat_region   31126..31164
                /rpt_family="LIME"
repeat_region   complement(31202..31562)
                /rpt_family="AT_rich"
repeat_region   complement(31609..31976)
                /rpt_family="THE1C"
repeat_region   31980..32021
                /rpt_family="MER57B"
repeat_region   32637..32797
                /rpt_family="(TG)n"
repeat_region   32905..32948
                /rpt_family="MIR"
repeat_region   complement(33347..33529)
                /rpt_family="CT-rich"
repeat_region   33642..33695
                /rpt_family="MER91A"
repeat_region   complement(34118..34426)
                /rpt_family="AT_rich"
repeat_region   34662..34971
                /rpt_family="AluSg"
repeat_region   34972..34996
                /rpt_family="AluSp"
repeat_region   36739..37053
                /rpt_family="(TAA)n"
repeat_region   37054..37093
                /rpt_family="AluJb"
repeat_region   37581..37607
                /rpt_family="(GAAA)n"
repeat_region   37647..37692
                /rpt_family="AT_rich"
repeat_region   37905..38174
                /rpt_family="AT_rich"
repeat_region   38229..38317
                /rpt_family="THE1B"
repeat_region   complement(38318..38683)
                /rpt_family="CT-rich"
repeat_region   38684..38708
                /rpt_family="MLT1B"
repeat_region   complement(38888..39013)
                /rpt_family="CT-rich"
repeat_region   39074..39186
                /rpt_family="MER5A"

Query Match   60.3%; Score 1022.2; DB 9; Length 146010;
Best Local Similarity   99.7%; Pred. No. 2e-240;
Matches 1024; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy   661 CAGATTCCAGGCTTCCTGATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGCCTTCAT 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 11578 CAGATTCCAGGCTTCCTGATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGCCTTCAT 11637

Qy   721 TATGACTGAATTACACTCATCCTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTTCACAGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 11638 TATGACTGAATTACACTCATCCTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTTCACAGC 11697

Qy   781 TCATCTTGCTGTGTCAATTATTACCATCACAGCCTTTAACAAAATCATCTTAAAATGC 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 11698 TCATCTTGCTGTGTCAATTATTACCATCACAGCCTTTAACAAAATCATCTTAAAATGC 11757

Qy   841 TACCCTTCAGCCTTAACCCTTAATGGAAAAATGAAAGGAAGTGACAATACGGGAGGTCCA 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 11758 TACCCTTCAGCCTTAACCCTTAATGGAAAAATGAAAGGAAGTGACAATACGGGAGGTCCA 11817

Qy   901 AACTTTGTCCCTGTTCTGTGTTCCTTACCTTTCGTCCCTGTGTATAGATTATGTAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 11818 AACTTTGTCCCTGTTCTCTGTTCCTTACCTTTCGTCCCTGTGTATAGATTATGTAAA 11877

```
Qy    961  AGCCTTGTGTAAATATGAGAGATGTTGTCAAAAATGATGCAGTAAATGATGAGCAATGACAGTGTA  1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11878  AGCCTTGTGTAAATATGAGAGATGTTGTCAAAAATGATGCAGTAAATGATGAGCAATGACAGTGTA  11937

Qy   1021  CTGCAGAGAAAATTTACTCCTTGCCTAGAACTGGGAGGGTTTTATGGGTCTGTAATTTTCC  1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11938  CTGCAGAGAAAATTTACTCCTTGCCTAGAACTGGGAGGGTTTTATGGGTCTGTAATTTTCC  11997

Qy   1081  CACACTCATTGCTGAAAGCTTAATTAAGTACTTCAAAAAACGTATCCCATTGTTTTACCT  1140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  11998  CACACTCATTGCTGAAAGCTTAATTAAGTACTTCAAAAAACGTATCCCATTGTTTTACCT  12057

Qy   1141  TCTTGAGGGAACGGTCTTGTTAACCAGCCCGAGTGTCTCACCCCAAACAATCCTCTGTC  1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  12058  TCTTGAGGGGAACGGTCTTGTTAACCAGCCCGAGTGTCTCACCCCAAACAATCCTCTGTC  12117

Qy   1201  ATTTTCAAAGATGCAAAAATGTGTTATTTAATTGTCTCCACCATTGTCACACACAGGAAT  1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  12118  ATTTTCAAAGATGCAAAAATGTGTTATTTAATTGTCTCCACCATTGTCACACACAGGAAT  12177

Qy   1261  GCCTAATAATAGCAACCCTTGTCTCCCTCTCTCTCCTTGCAAATGGCTCAGTGACTGG  1320
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  12178  GCCTAATAATAGCAACCCTTGTCTCCCTCTCTCTCCTTGCAAATGGCTCAGTGACTGG  12237

Qy   1321  AAGAGGCGGACTAATAGCCAGAGTTAAATATAAATACAAATTAATAATACATAGAGAACA  1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  12238  AAGAGGCGGACTAATAGCCAGAGTTAAATATAAATACAAATTAATAATACATAGAGAACA  12297

Qy   1381  GCAATACCAGAAAAAAGAATTCTGGTAAAATGATGTGAAAAATTGACAGCTCCCTCACT  1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  12298  GCAATACCAGAAAAAAAGAATTCTGTGAAAATGATGTGAAAAATTGACAGCTCCCTCACT  12357

Qy   1441  CTTAAGGTTGCTGCTATATACAGTCTAGGTTTCTGTTTGGAAATAGGTAGGGTAAAATC  1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  12358  CTTAAGGTTGCTGCTATATACAGTCTAGGTTTCTGTTTGGAAATAGGTAGGGTAAAATC  12417

Qy   1501  TAAGACCTGCACAAGGGCAGTGAGAGACATTTACAGCCCTCCTCTATTTGTTTTTTTAA  1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  12418  TAAGACCTGCACAAGGGCAGTGAGAGACATTTACAGCCCTCCTCTATTTGTTTTTTTAA  12477

Qy   1561  GGAAAAGTCAACTCCTGAAATGTCCCTTAGCTATAATCAGAAAAACTAAGAATATATTCT  1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  12478  GGAAAAGTCAACTCCTGAAATGTCCCTTAGCTATAATCAGAAAAACTAAGAATATATTCT  12537

Qy   1621  GTGTCAACAATGTATTTATGGAGAGAAGTAAAAATAAGTTCCACAGCAACACAAAAACAT  1680
           |||||||||||||||||||||||||||||||||||          ||||||||||||||
Db  12538  GTGTCAACAATGTATTTATGGAGAGAAGTAAAAATAAGTTCCACAGCAACACAAAAACAT  12597

Qy   1681  GAATTAT  1687
           |||||||
Db  12598  GAATTAT  12604
```

RESULT 9
BC050194

```
misc_feature    FSDRKRFGGGALKNFHCD"
                21..464
                /gene="Rab13"
                /note="RAB: Region: Rab subfamily of small GTPases"
                /db_xref="CDD:smart00175"
ORIGIN

Query Match     51.7%;  Score 875.6;  DB 10;  Length 2064;
Best Local Similarity  79.1%;  Pred. No. 3.2e-204;
Matches 1201;  Conservative  0;  Mismatches 279;  Indels 38;  Gaps 12;

Qy     17 AGATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTGTTGGGAAT  76
           ||||||||||||||||||||||||||||| ||||||||||||||||
Db      1 AGATGGCGTCCCTGGATCGGGTGAAGGTTCTGGTGTTGGGAAAT  60

Qy     77 CTTCGTTAGTCCATCTCCTATGCCAAATACAAGTGCTGGGAAATCATCATGGACTGTGG  136
           ||||||||| || || |||| |||||||| ||||||||||||||||||||||||||||
Db     61 CTTCACTCGTCCATCTCCTCGTGTGCCACAATCAAGTGTTAGGAAATCCGTCATGGACTGTGG  120

Qy    137 GCTGCCCAGTGGATGTCGAGAGTTCATGATTACAAAGAAGAAGGACCCCAGAAGAGAGACCT  196
           ||||||||| || ||||||||||||||||||||| ||||||||||||||||
Db    121 GCTGCCCGGTGGATATCAGAGTTCATGACTACAAAAGAAGAAGGGACCCCTGAAGAGAAGACAT  180

Qy    197 ACTACATAGAATTATGGGATGTTGGAGGCTCTGGGCCAGTGCCAGCAGCGGTGAAAAGCA  256
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACTATATAGAACTATGGGATGTTGGAGGCTCTGCTGGCCAGTGCCAGCAGTGTGAAAAGCA  240

Qy    257 CAAGAGCAGTATTCTACAACTCCGTAAATGTATTATTTCGTACACGACTTAACAAATA  316
           |||| |||||||||||||||||| |||||||| || || |||||||||||||||||
Db    241 CAAGAGCGGTGTCTACAACTCGTAAACGGCATCATTTTAGTACATAGACTACTTAACAAATA  300

Qy    317 AGAAGTCCTCCCAAAACTTGGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGC  376
           |||||||||||||||||||||||||||||||| ||||||||||||||||||| ||||
Db    301 AGAAGTCATCTCAAAACTTATATCGCGTGGTCTCAATGGAAGTTCTCAATCGGGATGCAGTTC  360

Qy    377 CAACTGGAGTCTCTTGGTGACAAATGGGATTATGATCAAGAGACAGTTGTCTGATAACCAAA  436
           |||||||||| ||| |||||||||||||||||||||||||||||||||| ||||||||
Db    361 CAACCGGAGTCCTGGTGACAAATGGGATTATGATCAACCAGAACAGTTGTCTGATAACCAAA  420

Qy    437 TACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGGCCATGAAG  496
           ||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCCCACTGTTGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGGCCATGAAG  480

Qy    497 TTTTAACTAGGACTGCTTCCTGGCTGAGGATTCAATCCAGAAGAGAAATTAATTTGGACT  556
           |||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    481 TTTTAATCAGGACTGCTTCTCTGGCTGAGGACTTCAATGCTGAAGAGATTAATTTGGATT  540

Qy    557 GCACAAATCCACGGTACTTAGCTGCAGTTCTTGCAGGTTCTCAGCTCGTCAGTAGGTTTT  616
           |||||||||||||||| |||||||||||||||||||||||||||||||||||
Db    541 GCACAAACCCACCGTGTACTTCCTGGCTGCAGTTCTCAGCTCGTCAGTAGATTTT  600

Qy    617 TTGATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAATCAGATTCCAGGCTTTC  676
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    601 TTGATAAGGTTATAGAGAAGAGATATTTTTTTAGAGAAGGTAATCAGATTCCGGGGCTTTT  660

Qy    677 CTGATCGGAAAAGATTTGGGCGCAGGAACATTAAAGAGCCTTCATTATGACTGAATTACAC  736
           ||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    661 CTGATCGGAAGAGAGGTTTGGAGGGAAGGGCACTGAACTTCCACTGTGACTGAGTGGCCC  720

Qy    737 TCATCCTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTTTACACGCTCATCTTGCTGTGTTC  796
           ||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    721 TCAGCC-CTGGAAGCAGCGCAACAGCAGTGTTCCACAGCTTCCTCTTGCTGTCGTGTAG  779

Qy    797 AATTATTACCATCACAGCCTTTTAACAAAATCATCTTAAAATGCTACCACCCTTCAGCCTTAC  856
           || || || || || ||||||     |||||||||||||||||||
Db    780 GGACATTAACTTCTCAGGC-TTTGGGAAAAACACATCTCAAAATGCTA-------CCTCAC  830

Qy    857 CCTTTAATGGAAAAATGAAAGGAAGTGACAATACGGAGGTTCCAAACTTTGTCCCTGTTC  916
           ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db    831 CCTCTCTGGAAAAATTGAAAGGAAGAGAGCCATGGAAGGTCCTAGCTCCTGCCCATTCTC  890

Qy    917 TCTGTTGTTCCTTACCTTCTGTCCCTGTGTATAGATTTATGTAAAAAGCCTTGTTGGTAAATAT  976
           | ||||| ||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCT  60
```

```
Db    891 TGTA---TCCTCACCACCCTGTCCCTGATTATAGCTTATGTAAAAGCCTTGTGACGACAT  947
           |||||||||||||||||||||||||||||||||||||||| || ||||||||||||||
Qy    977 GAGATGTTGTCAAAATGATGCAGTAAATGAGCAATGAGCAGTGTACTGACGAGAAAATTTA  1036
           |||||||||||||||| ||||||||||||||||| ||||||||||||||||||||
Db    948 AAAGGCTTGTCAAACCCATGCAGTAAAGAACTGGAGACA-AGGGCTGCAGCGAAGGCTGA  1006

Qy   1037 CTCTTGCCTAGACTGGAGGGGTTTTTATGGTCTGTAATTTTCCCACACTCATTGCTGAA  1096
           ||||||||||||||||||||||   ||| |||||||||||||||||||||||||||
Db   1007 CTCTTACTTGGAACCAGCGGG-TTTACAGGCCTGCATTTCCCACACTTGCTGCTGAA  1065

Qy   1097 AGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTACCTTCTTGAGGGAACGGT  1156
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1066 GGCTTAATTAAGTACTTCAGAAACATATCTGTGTTTTTA-CTTCCTGAGGGGAAAGGT  1124

Qy   1157 CTTGTTAACAGCCCGAGTTGTGTACCCCAAACAATCTCGTCATTTTCAAAGATGCAA  1216
           ||||||||||||||| ||| |||||||||||||||||||||||| |||||||||||
Db   1125 CATGTTAATCAGCCCGAGTTATCCACCCCCAAACCATCCCTGCCAATTTTAAAAGAAGCCA  1184

Qy   1217 AATGGTGTTATTTAATTGTCTCCACCATTGTCACACAGGAATGCCTAATAATAGCAAC  1276
           |||| || ||| |||| ||||||||||| || |||||||||||||||||||||||||
Db   1185 AATGGTGTTATTTCATTTTCTCCATCCTCATTACACACAGGGATGCCTAAGAATAGCAAC  1244

Qy   1277 CCTGTCTCCCCTCTCTCCCTTGCAAATGGTTCAGTCAGTCAGTCG---AAGAGGGGGACTA  1333
           ||     ||||||||||||||||||||||||||||||||     |||||||||||
Db   1245 CC--CATCTCCTGCTGCTCTCCTGGGAAAGGCGTTAGTGTCTGGGAAGAGCGATGGGACTA  1302

Qy   1334 ATAGCCAGAGTTAAAATATAAATACAAATTAATAAATATA------ACCAGAAA  1393
           |||||||||||||||||||||||||||||||||||      |||||||
Db   1303 ATAGCTGGAGTTAAAATAGAAATACAAATTAATAATAATCGTGTCTC-CAAGGTTGCAG  1358

Qy   1394 AAAAGAATTCTGTGTAAAATGATGTGAAAAATTGACAGCTCCCTCCACTCTTAAGGTTGCTG  1453
           | |||                 |||||||| |||||||
Db   1359 AGAAGGGT-------------CGTGAGACTCTAACAGCTCCTGCTC-CAAGGTTGCAG  1405

Qy   1454 CTATATACAGTCTAGGTTTCTGTTTGGAAATAGGTAGGGTAAAATCTTAAGACCTGCACA  1513
           |||||||||||||| ||||| |||||||||||||||||||| ||||||||||||||||
Db   1406 CTGTCGGAAGGGCTCTTTCTTGCTTTCTCGTTTTGGAAAGTGGGTGAAATCTGAGACCTGTACA  1465

Qy   1514 AGGGCAGTGAGAGACATT  1531
           |||| |||| |||| |||
Db   1466 AGGCCAGTAAGAGACAGATT  1483
```

```
RESULT 10
AX465671
LOCUS       AX465671               711 bp    DNA        linear    PAT 16-JUL-2002
DEFINITION  Sequence 3 from Patent WO0218425.
ACCESSION   AX465671
VERSION     AX465671.1  GI:21899931
KEYWORDS
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Meyers,R.A.
  TITLE     47324, a human g-protein and uses therefor
  JOURNAL   Patent: WO 0218425-A 3 07-MAR-2002;
            Millennium Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1..711
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match     42.0%;  Score 711;  DB 6;  Length 711;
Best Local Similarity  100.0%;  Pred. No. 9.9e-164;
Matches 711;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     19 ATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTGTTGGGAGACTCAGGTGTTGGGAAATCT  78
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCT  60
```

```
Qy    79 TCGTTAGTCCATCTCCTATGCCAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGGC 138
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TGCTTAGTCCATCTCCTATGCCAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGGC 120

Qy   139 TGCTCAGTGGATGTCAGAGTTCATGATGATTACAAAGAGGAACCCCAGAAGAGAGACCTAC 198
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TGCTCAGTGGATGTCAGAGTTCATGATGATTACAAAGAGGAACCCCAGAAGAGAGACCTAC 180

Qy   199 TACATAGAATTATGGGATGTTGGAGCTCTGTGGGCAGTGCCAGCAGCGGTGAAAAGCACA 258
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TACATAGAATTATGGGATGTTGGAGCTCTGTGGGCAGTGCCAGCAGCGGTGAAAAGCACA 240

Qy   259 AGAGCAGTATTCTTCACAACTCCGTAAATGGTATTATTTCGTACACGACTTAACAAATAAG 318
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AGAGCAGTATTCTTCACAACTCCGTAAATGGTATTATTTCGTACACGACTTAACAAATAAG 300

Qy   319 AAGTCCTCCCAAAACTGCCTCGTTGGTCGTTCGTTGGTCGTTGGTGAAGCTCTCAACAGGGATTTGGTGCCA 378
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AAGTCCTCCCAAAACTGCCTCGTTGGTCGTTCGTTGGTCGTTGGTGAAGCTCTCAACAGGGATTTGGTGCCA 360

Qy   379 ACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGAAACAGTTTGCTGATAACCAAATA 438
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 ACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGAAACAGTTTGCTGATAACCAAATA 420

Qy   439 CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAAGCGCCATGAAGTT 498
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAAGCGCCATGAAGTT 480

Qy   499 TTAACTAGGACTGCTTTCCTGCTGAGGATTCAATCCAGAGAAGAATTAATTTGGACTGC 558
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TTAACTAGGACTGCTTTCCTGCTGAGGATTCAATCCAGAGAAGAATTAATTTGGACTGC 540

Qy   559 ACAAATCCACGGTACTTAGCTCGCAGGTTCTCCAATGCTGTCAGCTCGTCAGTAGGTTTTT 618
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 ACAAATCCACGGTACTTAGCTCGCAGGTTCTCCAATGCTGTCAGCTCGTCAGTAGGTTTTT 600

Qy   619 GATAAGGTCATAGAGAAGAGATACTTTTAAGAGAAGGTAATCAGATTCAGATTCCAGGCTTCCT 678
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GATAAGGTCATAGAGAAGAGATACTTTTAAGAGAAGGTAATCAGATTCAGATTCCAGGCTTCCT 660

Qy   679 GATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGCCTTCATTATGACTGA 729
         |||||||||||||||||||||||||||||||||||||||||
Db   661 GATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGCCTTCATTATGACTGA 711
```

```
RESULT 11
LOCUS       AJ720020
AJ720020
DEFINITION  Gallus gallus mRNA for hypothetical protein, clone 9j2.
ACCESSION   AJ720020
VERSION     AJ720020.1  GI:53130699
KEYWORDS    ORF1.
SOURCE      Gallus gallus (chicken)
ORGANISM    Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1
AUTHORS     Caldwell,R.B., Kierzek,A.M., Arakawa,H., Bezzubov,Y., Zaim,J.,
            Fiedler,P., Kutter,S., Blagodatski,A., Kostovska,D., Koter,M.,
            Plachy,J., Carninci,P., Hayashizaki,Y. and Buerstedde,J.M.
TITLE       Full-length cDNAs from bursal lymphocytes to facilitate gene
            function analysis
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 3327)
AUTHORS     Caldwell,R.B.
TITLE       Direct Submission
JOURNAL     Submitted (20-MAY-2004) Caldwell R.B., GSF - Forschungszentrum,
            Institut fuer Molekulare Strahlenbiologie, Ingolstaedter Landstr.
            1, D-85764 Neuherberg, GERMANY
FEATURES             Location/Qualifiers
     source          1..3327
```

```
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="CB"
                     /db_xref="taxon:9031"
                     /clone="9j2"
                     /cell_type="lymphocyte"
                     /tissue_type="bursa"
                     /clone_lib="riken1"
                     /dev_stage="2 weeks old"
     gene            1..3327
                     /locus_tag="RCJMB04_9j2"
     5'UTR           <1..30
                     /locus_tag="RCJMB04_9j2"
     CDS             31..723
                     /locus_tag="RCJMB04_9j2"
                     /note="ORF1"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAG31679.1"
                     /db_xref="GI:53130700"
                     /translation="MAALDRVKVLVLGDSGVGKSSLVHLLCHNQVLGNPSWTVGCSVD
                     VRIHDYKEGTPEEKTYIELWDVGGSVGSATSVKSTRAVFYNSLNGIILVHDLTNKKS
                     SQNLYRWSLEALNRDVAPTGVLVTNGDYDREQFADNQIPLLVIGTKLDQIPETKRNEV
                     LTRTAPLAEDFNAEEINLDCTNPRYLAAGSSNAVKLSRFDKVIEKRYFFKRWQSRFL
                     ASRKERDLEEH"
     3'UTR           724..>3327
                     /locus_tag="RCJMB04_9j2"
ORIGIN
Query Match             28.8%;  Score 488.2;  DB 5;  Length 3327;
Best Local Similarity   81.9%;  Pred. No. 4.7e-109;
Matches  587;  Conservative    0;  Mismatches  128;  Indels    2;  Gaps    2;
```

```
Qy    16 AAGATGCGTCCCTGCGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGTGTTGGGAAA 75
         |||| ||| ||| |||||| || ||| |||||||||||||||||||||||||||||
Db    28 AAGATGCGGCCTCTGATAGGGTCAAGGTCGGTGTTGGGGGGACTCCGTGTGTCGGGAAAG 87

Qy    76 TCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACTGTG 135
         |||||||| ||| ||||| || ||||||| ||||| |||||| |||||||| ||||||
Db    88 TCTTCCGCTGCTGCACCTCCTGTGTCACCAGGTGCCCGTGCCTCGGGGATGTCCTGGACGTG 147

Qy   136 GGCTGCTCAGTGGATGTCAGAGTTCATGATGATTACAAAGAGGAACCCCAGAAGAGAGACC 195
         ||||||| || |||||||||||||| |||||||| ||||| ||||||||||||||||||
Db   148 GGCTGCTCGGTGGATGTCGGATGTCGGATGTGCGAATCATGACTACAAAGAAGAAGAAGACA 207

Qy   196 TACTACATAGAATTATGGGATGTTGGAGGTTCAATTAATT 255
         ||| |||||||||||||| ||| |||| || ||||| |||
Db   208 TACTATATAGAGCGTGTGGGATGTTGGAGGTTCAGTGGGTAGTGGTCCACTAGTGTGAAAAGC 267

Qy   256 ACAAGACAGTATTCTCACAACTCCGTAAATGTATTATTTCGTACACGACTTAACAAT 315
         ||||||||||||| |||||| ||||||||||| ||||| |||| ||| ||||||| ||
Db   268 ACAAGACAGTATTTTTATACCTCGCTGAATGGATAATTTTTAGTGCATGACTTAACCAAC 327

Qy   316 AAGAAGTCCTCCCAAAACTTGCGTCGTTGGTCGTTCATTGGAAGCTCTCAACAGGGATTGGTGTG 375
         ||||| |||||| |||| || || || || || ||| |||| |||||||| ||||||||| |||
Db   328 AAGAAATCATCCCAGAATTTATATCGCTGGTCTTTGGAAGCACTCAACAGACTGTGCGCT 387

Qy   376 CCAACTGGAGTCTTGGTGACAAATGGGGATTATGACAAATGGGGATTTGCTGATAACCAA 435
         |||||| |||| ||||||||| |||||||||||||| |||| |||| |||||||||||||
Db   388 CCCACAGGAGTTCTTGTGACAAAACGGTGACTACGACCGTGACAGTTTGCTGATAACCAG 447

Qy   436 ATACCACTGTTGGTAATAGGGACTAAACTGGACTAGATTCATGAAAAGCGCCCATGAA 495
         |||||||| ||||| ||||| ||| ||| |||||||| |||||||| ||| ||||||
Db   448 ATTCCACTGTTTATATAACTCGCTGAATGGGATAATTTTTAGTGCATGACTTAACCAAC 507

Qy   496 GTTTTAACTAGGACTGCTTCCTGGCTGAGGATTCAATCCAGAGAAGAAATTAATTTGGAC 555
         ||| |||||||| ||| || || || |||||| ||| ||| |||||||| |||||||| ||
Db   508 GTTTTGACCAGGACAGCTTTCTTGGCTGAGGATTCAATGCTGAGAGATAAATTTGGAT 567

Qy   556 TGCACAAATCCACGGTACTTAGCTCGCAGGTTCTTCCAATGCTGTCAGCTCGTCAGTAGGTTT 615
         ||||||| ||||| | ||||| || || ||| |||| ||| |||| || |||| ||||||| |
Db   568 TGCACCAATCCTCGTTACTTGGCGGCTCATCCAATGCTCATCCAATGCTGTCAAACTAAGCTCAAGAGATT 627
```

```
Qy    616  TTTGATAAGGTCATAGAGAAGAGATACTT-TTTAAGAGAAGTAATC-AGATTCCAGGCT 673
           ||||||| ||||||||||||||||||||| |||||||||||||| ||||||||||||||
Db    628  TTTGATAAGGTGATAGAGAAGAGATACTTCTTTAAGAGAAGATGGCAATCAAGATTCCTGCT 687

Qy    674  TCCTCGATCGGAAAAGATTTGGGGCAGGAACATTAAGAGAGCCTTCATTATGACTGAA 730
           | ||| |||||||||||||||||| |||||||  ||||||||||||||||||||||||
Db    688  TCTCGGAAGAAGAGATTTGAGGAGGACACTGAGAGAGCCTTCATTATGACTGAA 744
```

RESULT 12
BC073035

LOCUS   BC073035   1064 bp   mRNA   linear   VRT 03-AUG-2004
DEFINITION Xenopus laevis MGC82648 protein, mRNA (cDNA clone MGC:82648
            IMAGE:5049322), complete cds.
ACCESSION   BC073035
VERSION   BC073035.1 GI:49257300
KEYWORDS   MGC.
SOURCE   Xenopus laevis (African clawed frog)
  ORGANISM Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Xenopus.
REFERENCE   1 (bases 1 to 1064)
  AUTHORS   Klein,S.L., Strausberg,R.L., Wagner,L., Pontius,J., Clifton,S.W.
            and Richardson,P.
  TITLE   Genetic and genomic tools for Xenopus research: The NIH Xenopus
            initiative
  JOURNAL   Dev. Dyn. 225 (4), 384-391 (2002)
  PUBMED   12454917
REFERENCE   2 (bases 1 to 1064)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE   Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  PUBMED   12477932
REFERENCE   3 (bases 1 to 1064)
  AUTHORS   Klein,S. and Gerhard,D.S.
  TITLE   Direct Submission
  JOURNAL   Submitted (07-JUN-2004) National Institutes of Health, Xenopus Gene
            Collection (XGC), National Institute of Child Health and Human
            Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD
            20892-7510, USA
REMARK   NIH-MGC Project
COMMENT   Contact: XGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Igor Dawid
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,
            Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth
            Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
            Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,
            Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prabhu,

            Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska,
            Duane Smailus, Jeff Stott, Miranda Tsai, George Yang, Jacquie
            Schein, Asim Siddiqui, Rob Holt, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 152 Row: d Column: 10
            This clone was selected for full length sequencing because it
            passed the following selection criteria: similarity but not
            identity to protein.
FEATURES   Location/Qualifiers
     source   1..1064
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="MGC:82648 IMAGE:5049322"
                     /tissue_type="Ovary, Xenopus"
                     /clone_lib="NICHD_XGC_Ov1"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene   1..1064
                     /gene="MGC82648"
                     /db_xref="LocusID:444046"
     CDS   35..742
                     /gene="MGC82648"
                     /codon_start=1
                     /product="MGC82648 protein"
                     /protein_id="AAH73035.1"
                     /db_xref="GI:49257301"
                     /db_xref="LocusID:444046"
                     /translation="MASLDRVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVD
                     VRLHEYREGTPEKTYYTELWDVGGSVGSASSVKSTRAVFYNAVNGIILVHDLTNKKS
                     SQNLYRWSLEALNRDLQPMGVLVTNGDYDREQFADNQIPLLVIGTKLDQIPEAKRNEV
                     LTRTAFLABDFNAEEINLDCTNTRCLAAGSSNAVKLSRFFDKVIEKRYPREGNLIPGF
                     SDRKRFGGGNFKSLHYD"
ORIGIN
     Query Match   25.1%; Score 425; DB 5; Length 1064;
     Best Local Similarity 74.9%; Pred. No. 1.8e-93;
     Matches 546; Conservative 0; Mismatches 180; Indels 3; Gaps 1;
```

```
Qy      1  AGGGAAGGCAGTGCAAGATGGCCTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGAC   60
           ||||||| |||||||||| |||||||||||||||||||||||||||| ||||||||
Db     17  AGGTCTGACAGGGTAACATGGCCTTCCTGGACAGAGTGAAGGTTCTGGTGCTGGGGGAT   76

Qy     61  TCAGGTGTTGGGAAATCTTCGTTAGTCATCTCCTATGCCAAATCAAGTGCTGGGAAAT  120
           ||| ||| |||||||||||||||| |||||| |||||| |||||| |||||| ||||
Db     77  TCAGGAGTCGGGAAGTCTTCGTTGGTGCATCTGCTGTGTCAGAATCAGGTTTGGGAAC  136

Qy    121  CCATCATGGACTGTGGGCTGCCAGTGGATGTGCAGAGTTCAGAGTTCATGATTACAGAA  180
           ||||||||||||||||| |||||||||||||||||||| |||||||||| | |||||||
Db    137  CCCTCGTGGACAGTTGGCTGCCTGCAGTGAGTGTGCGGCTCCACGAGTACAGGGAGGGACT  196

Qy    181  CCAGAAGAGAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGGGCAGTGCC  240
           |||||||||| ||||||||||||||||||| || ||||||||||||| || |||||||
Db    197  CCAGAGGAGAAAACCTATTACACTGAGCTGTGGGACGTGGGGTCCGTTGGCAGTGCC  256

Qy    241  AGCAGCGGTGAAAAGCACAAGAGCACCAGGGCTGTGTTTTATAATGCAGTAAATGGGAT  300
           |||||||| |||||||||||| ||||| ||||| |||| |||||||||||| |||||||
Db    257  AGCAGTGTGAAGAGCACCAGGGCTGTGTTTTATAATGCAGTAAATGGGATCATCCTTGTT  316

Qy    301  CACGACTAACAAATAAGAAGTCCTCCCAAAACTTGCCTCGTCGTTGGTCATTGGAAGCTCTC  360
           ||| || ||||| ||||||||| || |||||||||||||| ||||||||||||||||||||||
Db    317  CATGACTTGACAAATAAGAAATCATCCAGAACCTTATCGCGTCCGTTGGTCATTGGAAGCCCTG  376

Qy    361  AACAGGGATTGGTGCCAACTGGAGTCTTGGTGCAAATGGGATTATGATCAAGAACAG  420
           |||||||||||||||| || || |||| |||||||| || ||||||||| || ||||
Db    377  AACCGAGATCTGCAGCCAATGGGAGTGCTAGTCACTAATGGGGACTACGACAGAGAGCAA  436

Qy    421  TTTGCTGATAACCAAATACCCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAA  480
           ||||||||||||||||||||| ||||||||||||||||||| || ||||||||||||||||
Db    437  TTTGCTGATAATCAGATCCCATTGTTGGTCATTGGAACCAAACTGGACCAAATTCCAGAG  496
```

```
Qy    481  ACAAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAA 540
            ||||||| |||| |||||| |||||| |||||||||||||||| ||| |||||||
Db    497  GCGAGACGCAATGAGGTCCTGACACGAACGAACTGCCTTTCTAGCTGAGGATTTTAATGCTGAA 556

Qy    541  GAAATTAATTGGACTGCACAAATCCACGTACTTAGCTGCCAGGTTCTTCCAATGCTGTC 600
            |||| |||||| ||||| |||||||||||||||| ||||| |||||||||||||||
Db    557  GAGATTAATCTTGACTGTACAACACGATGCCTGGCTGCTGGGTCTTCCAATGCTGTC 616

Qy    601  AAGCTCAGTAGGTTTTTGATAAGGTCATAGAGAGAAGAGATACTTTTTAAGAGAAGGTAAT 660
            |||||||||||| |||||||||||||||||||||||||||||||| ||||||||||||
Db    617  AAACTCAGCAGCAGTTTTTGACAAGGTTATTGAGAAAGATAC---CCACGGGAAGGAAAT 673

Qy    661  CAGATTCCAGCTTTCCTGATCGGCGAAAAGATTTGGGCGCAGGAACATTAAAGAGCCTTCAT 720
            |||||| || |||||| |||||||||||||||||||||||||||||||||||||| ||||
Db    674  CTGAGCCCTGGTTTCTTCCCGATCGGGAAACGCTTTGGAGGGGGGGAACCTTTAAGAGTCTCCAC 733

Qy    721  TATGACTGA 729
            |||||||||
Db    734  TATGACTGA 742
```

RESULT 13
```
BC078191
LOCUS        BC078191              1040 bp    mRNA    linear   VRT 26-JUL-2004
DEFINITION   Danio rerio cDNA clone MGC:100876 IMAGE:7141370, complete cds.
ACCESSION    BC078191
VERSION      BC078191.1  GI:50603855
KEYWORDS     MGC.
SOURCE       Danio rerio (zebrafish)
  ORGANISM   Danio rerio
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
             Cypriniformes; Cyprinidae; Danio.
REFERENCE    1  (bases 1 to 1040)
  AUTHORS    Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
             Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
             Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
             Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
             Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
             Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
             Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
             Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
             Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
             McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
             Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
             Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
             Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
             Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
             Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
             Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
             Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
             Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE      Generation and initial analysis of more than 15,000 full-length
             human and mouse cDNA sequences
  JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED    12477932
REFERENCE    2  (bases 1 to 1040)
  AUTHORS    Strausberg,R.
  TITLE      Direct Submission
  JOURNAL    Submitted (19-JUL-2004) National Institutes of Health, Mammalian
             Gene Collection (MGC), Cancer Genomics Office, National Cancer
             Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
             USA
  REMARK     NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      Contact: MGC help desk
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Genome Institute of Singapore
             cDNA Library Preparation: S. Mathavan, Chia-Lin Wei, and Yijun
             Ruan, Genome Institute of Singapore
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: National Institutes of Health Intramural
             Sequencing Center (NISC),
             Gaithersburg, Maryland;
```

```
             Web site: http://www.nisc.nih.gov/
             Contact: nisc_mgc@nhgri.nih.gov
             Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
             Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,P.,
             Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
             Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
             Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
             McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
             Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
             Young,A., Zhang,L.-H. and Green,E.D.

             Clone distribution: MGC clone distribution information can be found
             through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
             Series: IRAL Plate: 60 Row: e Column: 9
             This clone was selected for full length sequencing because it
             passed the following selection criteria: Hexamer frequency ORF
             analysis, Similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1..1040
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /strain="Singapore local strain"
                     /db_xref="taxon:7955"
                     /clone="MGC:100876 IMAGE:7141370"
                     /tissue_type="Embryo, 7 different stages (from just
                     fertilized embryos to 72 hours just hatched baby fish)"
                     /clone_lib="GIS2F001_ra"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
     CDS             42..743
                     /codon_start=1
                     /product="Unknown (protein for MGC:100876)"
                     /protein_id="AAH78191.1"
                     /db_xref="GI:50603856"
                     /translation="MASLDRVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVD
                     VRVHDYREGTPEEKAFYIELWDVGGSVGSASSVKSTRAVFYNSVNGIILVHDLTNKKS
                     SQNLYRWSLEALSKDSSPTGIIVSNGDYDREQFAENAVPLLLIGTKFDQIPENKRNDV
                     LTRTAPLSEDFNAEEINLDCTNPRYLAAGSNAVKLSRFFDKVIEKRYFTRDPSQMQS
                     FTDRRRFNFKSLHSD"
ORIGIN

Query Match             23.4%;  Score 397;  DB 5;  Length 1040;
Best Local Similarity   74.2%;  Pred. No. 1.3e-86;
Matches  502;  Conservative  0;  Mismatches 175;  Indels  0;  Gaps  0;
```

```
Qy     17  AGATGGCGTCCCTGGATCGGGTGAAGTACTGCTGTGTTGGGAGACTCAGGTGTTGGGAAAT 76
            |||||||| ||||||||| ||||||||||||||||||||||||||||||||| |||||
Db     40  AAATGGCCATCTTTGGACAGGGTGAAAGTGCTGGTTTTGGGAGATTCTGGAGTAGGGAAGT 99

Qy     77  CTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCATCATGACTGTGG 136
            |||||||||||||| |||| |||| ||||||| ||||||||||||||||||||||||
Db    100  CCTCTCTGTACACCTTCTTTGCCAGAATCAGTTTTGGGAATCCCTCATGGACTGTGG 159

Qy    137  GCTGCTCAGTGGATGTGCAGAGTTCATGATTACAAGAAGGAGACCCCAGAAGAAGAGACCT 196
            |||||||| |||||||||||||| |||||||||||||||||||||||||||||||||||
Db    160  GCTGCTCAGTGGACGTCAGGGTTCACGACTACAGAGAAGGCACTCCAGAAGAAGAGGCTT 219

Qy    197  ACTACATAGAATTATGGGACGTTGGAGTGTTGGAGGCTCGTGGGACTCAGGTGTTGGGAAAGCA 256
            |||||| ||||| |||||||||||| ||||||||||||||||||||||||||||| |||||
Db    220  TCTACATTGAACTCTGGGATGTTGGAGGATCTGTTGGCAGTGCTAGCAGTGTTAAAGCA 279

Qy    257  CAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTAACAAATA 316
            |||||| ||||| ||||| |||||||| || ||||||| ||| ||||||||||||||
Db    280  CCAGAGCTGTGTTTTACCAAAAACTGGTGTCAGTTTAATTTTAGTTTCACGATCGACTACA 339

Qy    317  AGAAGTCCTCCAAAACTGGCGTCGTTGGTCATTGGAAGCTTCAACAGGGATTTGGTGC 376
            |||||||||||||||| |||||| ||||||||||||| |||||||| |||||||||||
Db    340  AGAAGTCCTCCCAGAATCTATACCGCGTGGTCATTAGAGGCACTAAGCACTAAGCAAGAGACTCCTCTC 399

Qy    377  CAACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTAGTAACCAAA 436
            |||||||||||||||| ||| |||||||||||||||||||||||||||||||||||||
Db    400  CAACTGGCATCATCGATCATCGATCCAACGCGTGATTACGATTACGATTACGATGAGAGAAGAACGCTG 459
```

```
Qy      1 AAATGGCATCTTTGGACAGGGTGAAAGTGCGTGGTTTTGGGAGATTCTTGGAGTAGGGAGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     77 CTTTGGTTAGTCCATCTCCTATGCCAAAATCAACTGCTGGGGAAATCATCATGGACTGTGG 136

Qy     61 CCTCTTCTTGTACACCTTCTTTGCCGAGGATTTCAATCCAGAAGAAAATTAATTTGGACT 120
          |||||||| || ||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCTCTTCTTGTACACCTTCTTTGCCAGAATCAGTTTTGGGAATCCCTCCATGGACTGTGG 120

Qy    137 GCTGCTCAGTGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAAAGAAGAACCT 196
          |||||||||||||||||||||||||||| || ||||||||||||||||||||
Db    121 GCTGCTCAGTGATGTCAGGGTTCACGACTACCAGACTACGAGAGAGGCACTCCAGGAGAGGCTT 180

Qy    197 ACTACATAGAATTATGGACATGGGATGTTGGAGGCTTCTGTCGGGCAGTGCCAGCGCGTGAAAGCA 256
          ||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCTACATTGAACTCTGGGATGTTGGAGGATCTGTTGGCAGTGCCAGCAGTGTTAAAAGCA 240

Qy    257 CAAGAGCAGTATTCTACAACTCCGTAAATGTTATTATTTCTGTACACGACTTAACAAATA 316
          |||||||| |||||||||||||||||||||||||||||||||| ||||||||
Db    241 CCAGAGCTGTGTTTACAAATCTCAGTTACAATTCGTCAGTTTACCGAGAGCCGGATTTGGTGC 300

Qy    317 AGAAGTCCTCCCAAAACTTGCCGTCGTTGGTCATTGGAAGTCTCAACAGGAGATTTGGTGTGC 376
          |||||||||||||||| || ||| |||||||||||||||||||||||||
Db    301 AGAAGTCTTCCCAGAATCTATACCGCTGGTCACTCAGACGCACTGAGCAAGGACTCCTCTC 360

Qy    377 CAACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAAAGAGTTTGCTGATAACCAAA 436
          ||||||||||||||| ||||||||||||||||||||||||
Db    361 CAACTGGCATCATCGTATCCAACGGTGATTACGATAGAACAGTTTGCAGAGAATGCTG 420

Qy    437 TACCACTGTTGGTAATAGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAG 496
          |||||||||||||||||||||||||||| |||||||||
Db    421 TTCCTCTTCTGCTAATTGGCACCAAATTTGATCAGATCGATCCCGGAAAAACAAGAGAAGATG 480

Qy    497 TTTTAACTAGGACTGCTTCTCGGCTGAGGATTTCAATCCAGAAGACTTCTTCCAATGTGTCTCAGTAGGTTTT 556
          |||||||||||||||||||||||||||||||
Db    481 TCCTGACCCGAACTGCCTTTCTATCTCGAAGACTTCAATGCGGAAGAGATCAACCTGGATT 540

Qy    557 GCACAAATCCACGGTACTTAGCTGCTTTCTTCCAATGTGTCTCAGTAGGTTTT 616
          ||||||||||||||||||| |||||| || |||||||||||
Db    541 GCACAAATCCACGGTATCTTGCTGCTGTTCATCAAATGCTGTTAAACTGAGTCGATTCT 600

Qy    617 TTGATAAGGTCATGAGAAGATACTTTTAAGAGAAGGTAATCAGATTCAGATTCGGAATTCT 676
          |||||||||||||||||||||||||
Db    601 TTGATAAGGTAATAGAGAAGAAAAGATACTTCACAGAGACCCCTAGCCAGATGCGAGCTTCA 660

Qy    677 CTGATCGGAAAAGATTT 693
          ||| || |||
Db    661 CAGACAGGAGGCGCTTT 677
```

RESULT 14
AY391440

LOCUS   AY391440   1291 bp   mRNA   linear   VRT 07-OCT-2003
DEFINITION   Danio rerio hypothetical protein MGC23920-like protein mRNA,
                complete cds.
ACCESSION   AY391440
VERSION   AY391440.1 GI:37362247
KEYWORDS   .
SOURCE   Danio rerio (zebrafish)
  ORGANISM   Danio rerio
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
                Cypriniformes; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 1291)
  AUTHORS   Song,H.D., Wu,X.Y., Sun,X.J., Zhou,Y., Liu,T.X., Deng,M.,
                Zhang,G.W., Sheng,Y., Chen,Y., Ruan,Z., Jiang,C.L., Fan,H.Y.,
                Zon,L.I., KanKi,J.P., Look,A.T. and Chen,Z.
  TITLE   Gene Expression Profiling in the Zebrafish Kidney Marrow Tissue
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 1291)
  AUTHORS   Song,H.D., Wu,X.Y., Sun,X.J., Zhou,Y., Liu,T.X., Deng,M.,
                Zhang,G.W., Sheng,Y., Chen,Y., Ruan,Z., Jiang,C.L., Fan,H.Y.,
                Zon,L.I., KanKi,J.P., Look,A.T. and Chen,Z.
  TITLE   Direct Submission
  JOURNAL   Submitted (15-SEP-2003) State Key Lab for Medical Genomics,
                Shanghai Institute of Hematology, Ruijin Hospital Affiliated to
                Shanghai Second Medical University, 197 Rui Jin Road II, Shanghai
                200025, P. R. China
FEATURES   Location/Qualifiers
     source   1..1291
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /clone="RK025A2D04"
     CDS   3..704
                     /codon_start=1
                     /product="hypothetical protein MGC23920-like protein"
                     /protein_id="AAQ91252.1"
                     /db_xref="GI:37362248"
                     /translation="MASLDRVKVLVLGDSGVGKSSLVHLLCQNVLGNPSWTVGCSVD
                     VRVHDYREGTPEBKAFYIELWDVGGSVGSASSVKSTRAVFYNSVNGIILVHDLTNKKS
                     SQNLYRWSLEALSKDSSPTGIIVSNGDYDREQFAENAVPLLLIGTKFDQIPENKRNDV
                     LRTAFLSEDFNAEEINLDCTNPRYLAAGSSNAVKLSRFFDKVIEKRYFTRDPSQMQS
                     FTDRRRFNFKSLHSD"

ORIGIN

Query Match   23.3%; Score 395.4; DB 5; Length 1291;
Best Local Similarity 74.0%; Pred. No. 3.2e-86;
Matches 501; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

```
Qy    17 AGATGGCGTCCCTGGATCGGGTGAAGGTACTGGTCGTGGGAGACTCAGGTGTTGGGAAAT 76
         |||||| |||||| ||| ||||| || ||||| || ||||||||||||||||
```

```
Qy   437 TACCACTGTTGGTAATAGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAG 496
         |||||||||||||||||||||||||||| |||||||||
Db   460 TTCCTCTTCTGCTAATTGGCACCAAATTTGATCAGATCGATCCCGGAAAAACAAGAGAAGATG 519

Qy   497 TTTTAACTAGGACTGCTTCTCGGCTGAGGATTTCTATCTCGAAGACTTCAATGCGGAAGACTTCAGTAGGTTTT 556
         |||||||||| ||||||| |||||| | ||||||||||||||
Db   520 TCCTGACCCGAACTGCCTTTCTATCTCGAAGACTTCAATGCGGAAGAGATCAACCTGGATT 579

Qy   557 GCACAAATCCACGGTACTTAGCTGCTTTCTTCCAATGTGTCTCAGTAGGTTTT 616
         |||||||||||||||||||| |||||| || |||||||||||
Db   580 GTACAAATCCACGGTATCTTGCTGCTGTTCATCAAATGCTGTTAAATTGAGTCGATTCT 639

Qy   617 TTGATAAGGTCATGAGAAGATACTTTTAAGAGAAGGTAATCAGATTCAGATTCGGAATTCT 676
         |||||||||||||||||||||||||
Db   640 TTGATAAGGTAATAGAGAAGAAAAGATATTTTACAAGAGACCCCTAGCCAGATGCGAGCTTCA 699

Qy   677 CTGATCGGAAAAGATTT 693
         ||| || |||
Db   700 CAGACAGGAGGCGCTTT 716
```

RESULT 15
BD076273

LOCUS   BD076273   385 bp   DNA   linear   PAT 27-AUG-2002
DEFINITION   5' EST of tissue-nonspecific secretory protein.
ACCESSION   BD076273
VERSION   BD076273.1 GI:22621876
KEYWORDS   JP 2001512011-A/221.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 385)
  AUTHORS   Edwards,J.B.D.M., Duclert,A. and Lacroix,B.
  TITLE   5' EST of tissue-nonspecific secretory protein
  JOURNAL   Patent: JP 2001512011-A 221 21-AUG-2001;
                GENSET
COMMENT   OS   Homo sapiens (human)
                PN   JP 2001512011-A/221
                PD   21-AUG-2001
                PF   31-JUL-1998 JP 2000505289
                PR   01-AUG-1997 US   08/905135
                PI   JEAN BAPTISTE DUMAS MILNE EDWARDS,AYMERIC DUCLERT,BRUNO PI
                LACROIX
                PC   C12N15/09,C12N15/09,C07K14/47,C12Q1/68,C12N15/00,C12N15/00 CC
                blastn
                CC   identity 97

```
CC        region 1..347
CC        id AA023764
CC        est
CC        blastn
CC        identity 95
CC        region 145..384
CC        id CO3036
CC        est
CC        blastn
CC        identity 98
CC        region 2..71
CC        id CO3036
CC        est
CC        blastn
CC        identity 99
CC        region 1..193
CC        id R08519
CC        est
CC        blastn
CC        identity 94
CC        region 193..263
CC        id R08519
CC        est
CC        Von Heijne matrix
CC        score 4.8
CC        seq SLVHLLCQNQVLG/NP
CC        n=a, g, c or t
FH   Key            Location/Qualifiers
FT   misc_feature    39..385
FT   misc_feature    146..385
FT   misc_feature    11..80
FT   misc_feature    39..231
FT   misc_feature    232..302
FT   sig_peptide     11..109
FT   misc_feature    287
FT   misc_feature    331..

FEATURES             Location/Qualifiers
     source          1..385
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"

ORIGIN

Query Match     22.1%;  Score 374.8;  DB 6;  Length 385;
Best Local Similarity   97.4%;  Pred. No. 4.2e-81;
Matches  374;  Conservative   5;  Mismatches   5;  Indels   0;  Gaps   0;

Qy      10 AGTGGCAAGATGGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTT  69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 AGTGGCAAGATGGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTT  61

Qy      70 GGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGG 129
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 GGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGG 121

Qy     130 ACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAG 189
           ||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db     122 ACTGTGGGCTGCTCAGTGGATGTCAGAGTKCATGATTACAAAGAAGGAACCCCAGAAGAG 181

Qy     190 AAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTG 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 AAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTG 241

Qy     250 AAAAGCACAAGAGCAGTATTCTACACTCCGTAAATGGTATTATTTTCGTACACGACTTA 309
           |||||||||||||||||||||||||||||||||||::|||||||| :||||||||||||
Db     242 AAAAGCACAAGAGCAGTATTCTACACTCCGTAAATGGTATTATTWNYCGTACACGACTTA 301

Qy     310 ACAAATAACGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGAT 369
           ||:: ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 ACVSATGGGAAGTCCTCCCAAAAMTGCGNCGTTGGTCATTGGAAGCTCTCTCAACAGGGAT 361

Qy     370 TTGGTGCCAACTGGAGTCTTGGTG 393
           ||||||||||||||||||||||||
Db     362 TTGGTGCCAACTGGAGTCTTGGTG 385
```

Search completed: June 30, 2005, 23:29:32
Job time : 7774.11 secs

```
                  GenCore version 5.1.6
          Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:       June 30, 2005, 13:26:53 ; Search time 1017.81 Seconds
                                        (without alignments)
                                        9852.580 Million cell updates/sec

Title:        US-09-945-173-1
Perfect score: 1694
Sequence:     1 agggaaggcagtggcaagat.........aaacatgatgattattgaacta 1694

Scoring table: IDENTITY_NUC
              Gapop 10.0 , Gapext 1.0

Searched:     4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters:        8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    N_Geneseq_16Dec04:*
         1:   geneseqn1980s:*
         2:   geneseqn1990s:*
         3:   geneseqn2000s:*
         4:   geneseqn2001as:*
         5:   geneseqn2001bs:*
         6:   geneseqn2002as:*
         7:   geneseqn2002bs:*
         8:   geneseqn2003as:*
         9:   geneseqn2003bs:*
        10:   geneseqn2003cs:*
        11:   geneseqn2003ds:*
        12:   geneseqn2004as:*
        13:   geneseqn2004bs:*

                         SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

Result No.	Score	Query % Match	Length	DB	ID	Description
1	1692.4	99.9	1694	6	AAD34049	Aad34049 Human G-p
2	1664.6	98.3	2398	10	AAD54288	Aad54288 Human SEC
3	1358.8	80.2	1417	8	ABX63467	Abx63467 Human cDN
4	735	43.4	735	10	ADE47763	Ade47763 Human NOV
5	735	43.4	735	12	ADJ79033	Adj79033 Human NOV
6	731.6	43.2	739	10	ADE47765	Ade47765 Human NOV
7	731.6	43.2	739	12	ADJ79035	Adj79035 Human NOV
8	628.4	37.1	959	4	AAS27286	Aas27286 cDNA enco
9	628.4	37.1	959	10	ADB93464	Adb93464 Human cDN
10	444	26.2	504	9	ACH36304	Ach36304 Human end
11	374.8	22.1	385	2	AAX41283	Aax41283 Human sec
12	321	18.9	321	6	AAC17946	Aac17946 Human sec
13	252.8	14.9	387	8	ABX52440	Abx52440 Bovine ES
14	66.8	3.9	919	4	ABL08307	Abl08307 Drosophil
15	60	3.5	60	6	ABN37775	Abn37775 Human spl
16	56.8	3.4	3048	6	ABL08306	Abl08306 Drosophil
17	56.8	3.4	3215	4	ABL14880	Abl14880 Drosophil
18	56.8	3.4	3888	4	ABL14882	Abl14882 Drosophil
19	44.2	2.6	1234	3	AAC59835	Aac59835 Human sec
20	44.2	2.6	11922	3	AAA70187	Aaa70187 Plasmodiu

21		2.6	6713	6	ABL33081	Abl33081 Human imm
22		2.6	6713	6	ABL34527	Abl34527 Human met
23		2.6	6713	6	ABL70252	Abl70252 Chemicall
24		2.6	6713	7	ADS99788	Ads99788 Complemen
25		43.8	751	8	AAI95038	Aai95038 Human neu
26	c	43.8	2000	8	ADA71938	Ada71938 Rice gene
27	c	43.4	5641	6	ABL33396	Abl33396 Human imm
28	c	42.4	444	5	ADL43671	Adl43671 Human ova
29		42	9789	2	AAT41852	Aat41852 cDNA enco
30	c	42	56153	4	AAS46793	Aas46793 Tumour su
31		41.2	2239	4	AAH55047	Aah55047 S. epider
32	c	41.2	2259	4	AAH54714	Aah54714 S. epider
33	c	41.2	3475	4	AAH54668	Aah54668 S. epider
34	c	41.2	13131	6	ABL92248	Abl92248 Chemicall
35	c	41.2	56153	4	AAS46794	Aas46794 Tumour su
36		41	2000	8	ADA71938	Ada71938 Rice gene
37		40.8	457	7	ACH49768	Ach49768 Human leu
38	c	40.8	495	13	ACN49798	Acn49798 Cotton pr
39	c	40.8	176001	12	ADK43203	Adk43203 Human pro
40	c	40.8	186739	12	ADK43195	Adk43195 Human pro
41		40.6	7305	4	AAK90435	Aak90435 Human dig
42		40.6	7305	5	AAD16679	Aad16679 Human pan
43	c	40.4	120239	13	ABD33244	Abd33244 Murine ca
44	c	40.2	5461	6	ABL32999	Abl32999 Human imm
45	c	40.2	5461	6	ABQ66978	Abq66978 Human ang

```
                                    ALIGNMENTS

RESULT 1
AAD34049
ID   AAD34049 standard; cDNA; 1694 BP.
XX
AC   AAD34049;
XX
DT   16-JUL-2002  (first entry)
XX
DE   Human G-protein (47324) cDNA.
XX
KW   Human; G-protein; 47324; cancer; signal transduction; tumour; apoptosis;
KW   cardiovascular disorder; liver; autoimmune; blood vessel; bone; hormonal;
KW   reproductive; fertility; brain; haematopoietic; metabolic; gene therapy;
KW   hypertension; atherosclerosis; Gaucher's disease; degenerative disease;
KW   multiple sclerosis; glycogen storage disease; obesity; acute meningitis;
KW   anorexia; cerebral oedema; pain; Alzheimer's disease; diabetes mellitus;
KW   Parkinson's disease; rheumatoid arthritis; osteoporosis; dermatological;
KW   transgenic animal; antiinflammatory; antithyroid; anorectic; vasotropic;
KW   anti-HIV; nootropic; vulnerary; ophthalmological; virucide; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   5'UTR          1..18
FT                  /*tag= a
FT   CDS            19..729
FT                  /*tag= b
FT                  /product= "Human 47324 protein"
PF                  /note= "CDS referred to as SEQ ID NO:3 is specifically
                    claimed in claim 1 of the specification"
FT   3'UTR          730..1694
FT                  /*tag= c
XX
PN   WO200218425-A2.
XX
PD   07-MAR-2002.
XX
PF   31-AUG-2001; 2001WO-US027197.
XX
PR   01-SEP-2000; 2000US-0229293P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
```

```
PI   Meyers RA;
XX
DR   WPI; 2002-315535/35.
DR   P-PSDB; AAE21568.
XX
PT   A human G protein and nucleic acid molecule encoding it, for diagnosing
PT   and treating cellular proliferative, bone, immune, cardiovascular, liver,
PT   pain or metabolic disorders and identifying modulators.
XX
PS   Claim 1; Fig 1; 118pp; English.
XX
CC   The invention relates to a human G-protein, termed 47324 and nucleic acid
CC   encoding it. 47324 is useful for treating cancer or aberrant cellular
CC   proliferation and/or differentiation such as breast, ovarian, prostate,
CC   colon and lung cancer, pain, heart disorders and viral disease. 47324
CC   associate a lipid bilayer due to post-translational modifications,
CC   modulate signal transduction, control cell morphology and
CC   differentiation, mediate cellular proliferation, apoptosis, modulate
CC   tumour inhibition, regulate cellular trafficking and active cytoskeleton
CC   organisation. Cardiovascular disorders include hypertension and
CC   atherosclerosis; liver disorders include Gaucher's disease and glycogen
CC   storage disease; metabolic disorders include obesity and anorexia; brain
CC   disorders include cerebral oedema and neural tube defect; infections
CC   include acute meningitis and multiple sclerosis; degenerative diseases
CC   include Alzheimer's disease and Parkinson's disease; blood vessel
CC   disorders include responses of vascular cell walls to injury;
CC   haematopoietic disorders include autoimmune disease such as diabetes
CC   mellitus and rheumatoid arthritis; bone disorders include osteoporosis;
CC   hormonal disorders include reproductive or fertility disorders. Ras
CC   family-associated or related disorders also include immune disorders such
CC   as autoimmune disorders or immune deficiency disorders e.g. severe
CC   combined immunodeficiency. 47324 is also useful as pharmacogenomic
CC   markers. 47324 is useful for producing non-human transgenic animals which
CC   are useful for studying the function and/or activity of 47324 and for
CC   identifying and/or evaluating modulators of 47324 activity. 47324 gene is
CC   useful in gene therapy. The present sequence is human 47324 cDNA
XX
SQ   Sequence 1694 BP; 530 A; 323 C; 360 G; 481 T; 0 U; 0 Other;

Query Match       99.9%;   Score 1692.4;   DB 6;   Length 1694;
Best Local Similarity   99.9%;   Pred. No. 0;
Matches  1693;   Conservative   0;   Mismatches   0;   Indels   1;   Gaps   0;

Qy     1 AGGGAAGGCAGTGGCAAGATGGCGTCCCTGGATCCGGTGAAGGTACTGCGTGTTGGGAGAC   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 AGGGAAGGCAGTGGCAAGATGGCGTCCCTGGATCCGGTGAAGGTACTGCGTGTTGGGAGAC   60

Qy    61 TCAGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAAT  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TCAGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAAT  120

Qy   121 CCATCATGGACTGTGGCTGTCAGTGGATGTCAGAGTTCAGATTCATGATTACAAAGAAGAACC  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CCATCATGGACTGTGGCTGTCAGTGGATGTCAGAGTTCAGATTCATGATTACAAAGAAGAACC  180

Qy   181 CCAGAAGAGAGAAGCACAAGAGACCTACTACATAGAATTATGGGATGTGGAGGCTCGTGGGCCAGTGCC  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CCAGAAGAGAGAAGCACAAGAGACCTACTACATAGAATTATGGGATGTGGAGGCTCGTGGGCCAGTGCC  240

Qy   241 AGCAGCGTGAAAAGCACAAGACCAACAGAGACCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTA  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AGCAGCGTGAAAAGCACAAGACCAACAGAGACCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTA  300

Qy   301 CACGACTTAACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTCGTTGGTCATTGGAAGCTCTC  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CACGACTTAACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTCGTTGGTCATTGGAAGCTCTC  360

Qy   361 AACAGGGATTTGGTGCCAACTGGAGTCTTCGTGCAACTGGGATTATGATCAAGAACAG  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AACAGGGATTTGGTGCCAACTGGAGTCTTCGTGCAACTGGGATTATGATCAAGAACAG  420

Qy   421 TTTGCTGATAACCAAATACCACTGTTGGTAATAGGACTAAACTGGACCAGATTCATGAA  480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 TTTGCTGATAACCAAATACCACTGTTGGTAATAGGACTAAACTGGACCAGATTCATGAA  480

Qy   481 ACAAAGCGCCATGAGAGTTTTAACTAGGACTGCTTCCTCGGCTGCTTTCCTCGGCTGCTTTCCAATGCTGTC  540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ACAAAGCGCCATGAGAGTTTTAACTAGGACTGCTTCCTCGGCTGCTTTCCTCGGCTGCTTTCCAATGCTGTC  540

Qy   541 GAAATTAATTGGACTGCACAAATCCACGGTACTTAGCTGCGAGGTTCTTCCAGGTTCTTCCAATGCTGTC  600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GAAATTAATTGGACTGCACAAATCCACGGTACTTAGCTGCGAGGTTCTTCCAGGTTCTTCCAATGCTGTC  600

Qy   601 AAGCTCAGTAGGTCATGATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAAT  660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 AAGCTCAGTAGGTCATGATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAAT  660

Qy   661 CAGATTCCAGGCTTCCTGATCGGAAAAGATTGGGCAGGAACATTAAAAGAGCCTTCAT  720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CAGATTCCAGGCTTCCTGATCGGAAAAGATTGGGCAGGAACATTAAAAGAGCCTTCAT  720

Qy   721 TATGACTGAATTACACTCACTCCTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTCACAGC  780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 TATGACTGAATTACACTCACTCCTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTCACAGC  780

Qy   781 TCATCTTGCTGTGTTCAATTATTACCATCACAGCCTTTAACAAAATCATCTTAAAATGC  840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 TCATCTTGCTGTGTTCAATTATTACCATCACAGCCTTTAACAAAATCATCTTAAAATGC  840

Qy   841 TACCCTTCAGCCTTACCCTTTAATGAAAAATGAAAGGAAGTGACAATACGGAGGTCCA  900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 TACCCTTCAGCCTTACCCTTTAATGAAAAATGAAAGGAAGTGACAATACGGAGGTCCA  900

Qy   901 AACTTTGTCCCGTTCTCGTGTTCCTTACCTTCGTCCCTGTGTATAGATTATGTAAA  960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 AACTTTGTCCCGTTCTCGTGTTCCTTACCTTCGTCCCTGTGTATAGATTATGTAAA  960

Qy   961 AGCCTTGTGTAAATATGAGATGTTGTCAAAATGATGAGCAGTAAATGAGCAATGACAGTGTA 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 AGCCTTGTGTAAATATGAGATGTTGTCAAAATGATGAGCAGTAAATGAGCAATGACAGTGTA 1020

Qy  1021 CTGCAGAGAAAATTTACTCTTGCCTAGAACTGGAGGGTTTTATGGGTCTGTAATTTCC 1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 CTGCAGAGAAAATTTACTCTGCCTAGAACTGGAGGGTTTTATGGGTCTGTAATTTCC 1080

Qy  1081 CACACTCATTGCTGAAAGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTTACCT 1140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 CACACTCATTGCTGAAAGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTTACCT 1140

Qy  1141 TCTTGAGGGGAACGGTCTTGTTAACCAGCCTGAGTGTCTACCCCAAACAATCTCTGTC 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 TCTTGAGGGGAACGGTCTTGTTAACCAGCCCTGAGTGTCTACCCCAAACAATCTCTGTC 1200

Qy  1201 ATTTTCAAAGATGCAAAATGGTGTTATTTAATTGTCTCACCACCAGGAAT 1260
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 ATTTTCAAAGATGCAAAATGGTGTTATTTAATTGTCTCACCACCAGGAAT 1260

Qy  1261 GCCTAATAATAGCAACCCTGTCTCCCTTGCAAATGGCTCAGTGACTGG 1320
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 GCCTAATAATAGCAACCCTGTCTCCCTTGCAAATGGCTCAGTGACTGG 1320

Qy  1321 AAGAGGCGGACTAATAGCCAGAGTTAAATATAATTAAATAATACATAGAGAACA 1380
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 AAGAGGCGGACTAATAGCCAGAGTTAAATATAATTAAATAATACATAGAGAACA 1380

Qy  1381 GCAATACCAGAAAAAAAGAATTCTGTAAAAATTGACAGCTCCCTCACT 1440
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 GCAATACCAGAAAAAAAGAATTCTGTAAAAATTGACAGCTCCCTCACT 1440

Qy  1441 CTTAAGGTTGCTACTATATACAGTTCTAGTTTTCTGTTTGGAAAATAGTAGGGTAAAATC 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 CTTAAGGTTGCTACTATATACAGTTCTAGTTTTCTGTTTGGAAAATAGTAGGGTAAAATC 1500

Qy  1501 TAAGACCTGCACAAGGGCAGTGAGAGACATTTACAGCCTCCTCTCTATTTGTTTTTTAA 1560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 TAAGACCTGCACAAGGGCAGTGAGAGACATTTACAGCCTCCTCTCTATTTGTTTTTTAA 1560
```

```
Qy    1561 GGAAAAGTCAACTCCTGAAATGTCCCTTAGCTATAATCAGAGAAAACTAAGAATATTATTCT 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 GGAAAAGTCAACTCCTGAAATGTCCCTTAGCTATAATCAGAGAAAACTAAGAATATTATTCT 1620

Qy    1621 GTGTCAACAATGTATTTATGGAGAGAAGTAAAAATAAGTTCCACAGCAACACAAAAACAT 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 GTGTCAACAATGTATTTATGGAGAGAGTATTAAAAATAAGTTCCACAGCAACACAAAAACAT 1680

Qy    1681 GAATTATTGAACTA 1694
           ||||||||||||||
Db    1681 GAATTATTGAACTA 1694
```

```
RESULT 2
AAD54288
ID    AAD54288    standard; cDNA; 2398 BP.
XX
AC    AAD54288;
XX
DT    17-JUN-2003 (first entry)
XX
DE    Human SECP-2 cDNA.
XX
KW    Human; secreted protein; SECP; cell proliferative disorder; cirrhosis;
KW    cancer; arteriosclerosis; hepatitis; psoriasis; developmental disorder;
KW    atherosclerosis; seizure disorder; renal tubular acidosis; anaemia;
KW    cataract; sensorineural hearing loss; neurological disorder; epilepsy;
KW    ischaemic cerebrovascular disease; Alzheimer's disease; Pick's disease;
KW    Huntington's disease; stroke; dementia; Parkinson's disease; anxiety;
KW    amyotrophic lateral sclerosis; schizophrenic disorder; mental disorder;
KW    Tourette's disorder; muscular dystrophy; autoimmune disorder; mood; AIDS;
KW    inflammatory disorder; acquired immunodeficiency syndrome; allergy; ARDS;
KW    adult respiratory distress syndrome; diabetes mellitus; Crohn's disease;
KW    asthma; autoimmune thyroiditis; glomerulonephritis; rheumatoid arthritis;
KW    atopic dermatitis; ulcerative colitis; trauma; cardiovascular disorder;
KW    infection; congestive heart failure; heart disease; angina pectoris;
KW    myocardial infarction; myocarditis; transgenic animal; gene therapy;
KW    transgenic; vaccine; virucide; antibacterial; fungicide; antiparasitic;
KW    protozoacide; gene; ss.
XX
OS    Homo sapiens.
XX
FH    Key            Location/Qualifiers
FT    CDS            3..713
FT                   /tag= a
FT                   /product= "Human SECP-2 protein"
FT    sig_peptide    3..101
FT                   /tag= b
FT    mat_peptide    102..710
FT                   /tag= c
FT                   /product= "Human mature SECP-2 protein"
XX
PN    WO200297035-A2.
XX
PD    05-DEC-2002.
XX
PF    21-MAY-2002; 2002WO-US016234.
XX
PR    25-MAY-2001; 2001US-0293728P.
PR    08-JUN-2001; 2001US-0297019P.
PR    19-JUN-2001; 2001US-0299297P.
PR    22-JUN-2001; 2001US-0300537P.
PR    29-JUN-2001; 2001US-0301936P.
PR    06-MAR-2002; 2002US-0362439P.
PR    08-MAR-2002; 2002US-0363649P.
PR    19-MAR-2002; 2002US-0366041P.
XX
PA    (INCY-) INCYTE GENOMICS INC.
XX
PI    Yue H, Lee EA, Becha SD, Baughn MR, Yao MG, Tang YT;
PI    Au-Young JK, Lal PG, Warren BA, Duggan BM, Tran UK, Xu Y;
PI    Thangavelu K, Richardson TW, Bandman O, Jones KA, Yang J;
```

```
PI    Emerling BM, Swarnakar A, Luo W, Walia NK, Azimzai Y, Khan FA;
PI    Lu DAM, Griffin JA, Lee SY, Burford N, Elliott VS, Honchell CD;
PI    He A, Mason PM, Li JX, Hafalia AJA, Gururajan R;
XX
DR    WPI; 2003-129519/12.
DR    P-PSDB; AAB35750.
XX
PT    Novel human secreted proteins and genes encoding the proteins, useful for
PT    treating, diagnosing and preventing cell proliferative,
PT    autoimmune/inflammatory, cardiovascular, developmental or neurological
PT    disorders.
XX
PS    Claim 89; Page 171-172; 192pp; English.
XX
CC    The present invention relates to novel human secreted proteins (SECP) and
CC    polynucleotides encoding such proteins. SECP sequences are useful for
CC    diagnosing, treating and preventing cell proliferative disorders
CC    including cancer (e.g. arteriosclerosis, cirrhosis, hepatitis, psoriasis
CC    and atherosclerosis), developmental disorders (e.g. seizure disorders,
CC    renal tubular acidosis, anaemia, cataract and sensorineural hearing
CC    loss), neurological disorders (e.g. epilepsy, ischaemic cerebrovascular
CC    disease, Alzheimer's disease, Pick's disease, Huntington's disease,
CC    stroke, dementia, Parkinson's disease, amyotrophic lateral sclerosis,
CC    schizophrenic disorders, mental disorders including mood and anxiety,
CC    Tourette's disorder and muscular dystrophy), autoimmune/inflammatory
CC    disorders (e.g. acquired immunodeficiency syndrome (AIDS), allergy, adult
CC    respiratory distress syndrome (ARDS), asthma, autoimmune thyroiditis,
CC    diabetes mellitus, Crohn's disease, glomerulonephritis, rheumatoid
CC    arthritis, atopic dermatitis, ulcerative colitis, trauma and viral,
CC    bacterial, fungal, parasitic, protozoal and helminthic infections) and
CC    cardiovascular disorders (e.g. congestive heart failure, ischaemic heart
CC    disease, angina pectoris, myocardial infarction, hypertensive heart
CC    disease, congenital heart disease and myocarditis). They are useful for
CC    creating knocking humanised animals or transgenic animals to model human
CC    diseases. Sequences of the invention are useful in somatic or germline
CC    gene therapy and in diagnostic purposes. They are also used as vaccines.
CC    The present sequence is human SECP-2 cDNA.
XX
SQ    Sequence 2398 BP; 722 A; 456 C; 470 G; 750 T; 0 U; 0 Other;

Query Match             98.3%;  Score 1664.6;  DB 10;  Length 2398;
Best Local Similarity   99.8%;  Pred. No. 0;
Matches 1667;  Conservative 0;  Mismatches 0;  Indels 4;  Gaps 0;

Qy      17 AGATGGGCGTCCCCTGAGGTGAAGGTACTGGTGTTGGGAGACTCAGTGTTGGGAAAT 76
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AGATGGGCGTCCCCTGAGGTGAAGGTACTGGTGTTGGGAGACTCAGTGTTGGGAAAT 60

Qy      77 CTTCGTTAGTCCATCTCCTATGCCAAAATCAACTCAGTGCTGGGAAATCCATCATGGACTGTGG 136
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 CTTCGTTAGTCCATCTCCTATGCCAAAATCAACTCAGTGCTGGGAAATCCATCATGGACTGTGG 120

Qy     137 GCTGCTCAGTGGATGTCAGAGTTCATGAGTTCATGATTACAAAGAAGAGGAACCCCAGAGAAGAAGACCT 196
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCTGCTCAGTGGATGTCAGAGTTCATGAGTTCATGATTACAAAGAAGAGGAACCCCAGAGAAGAAGACCT 180

Qy     197 ACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGAAAAGCA 256
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGAAAAGCA 240

Qy     257 CAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGGACTTAACAAATA 316
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGGACTTAACAAATA 300

Qy     317 AGAAGTCCTCCCAAAACTTGCGTCGTTGCTCATTGGAAGCTCTCAACAGGGATTTGGTGC 376
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 AGAAGTCCTCCCAAAACTTGCGTCGTTGCTCATTGGAAGCTCTCAACAGGGATTTGGTGC 360

Qy     377 CAACTGGAGTCGTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGATAACCAAA 436
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CAACTGGAGTCGTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGATAACCAAA 420

Qy     437 TACCACTGTTGGTAATAGGGACTAAACTCGACCAGATTCATGAAACAAAGCGCCATGAAG 496
```

```
Db  1501 GCAGTGAGAGACATTTACAGCCTCCTCTCTCATTTGTTTTTTTAAGGAAAAGTCAACTCCT 1560
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1577 GAAATGTCCCTTAGCTATATCAGAAAACTAAGAATATTATTCTGTGTCAACAATGTATT  1636
         |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db  1561 GAAATGTCCCTTAGCTATATCAGAAAACTAAGAATATTATTCTGTGTCAACAATGTATT  1620

Qy  1637 TATGGAGAAGTAAAATAAGTTCCACAGCAACACAAAACATGAATTAT             1687
         ||||||||||||||||||||||||||||||||||       ||||||||
Db  1621 TATGGAGAAGTAAAAATAAGTTCCACAGCAACACATTTACATGAATTAT           1671
```

RESULT 3
ABX63467
ID   ABX63467 standard; cDNA; 1417 BP.
XX
AC   ABX63467;
XX
DT   26-FEB-2003 (first entry)
XX
DE   Human cDNA #467 differentially expressed in activated vascular tissue.
XX
KW   Human; gene; ss; vascular tissue; cytostatic; atherosclerosis; cardiant;
KW   hypotensive; antidiabetic; gynaecological; vasotropic; cerebroprotective;
KW   gene therapy; vascular disease; cancer; coronary; artery disease;
KW   hypertension; diabetes; pre-eclampsia; restenosis;
KW   ischaemia-reperfusion injury; stroke.
XX
OS   Homo sapiens.
XX
PN   US2002137081-A1.
XX
PD   26-SEP-2002.
XX
PF   08-JAN-2002; 2002US-00044090.
XX
PR   28-JUL-2000; 2000US-0222469P.
PR   08-JAN-2001; 2001US-0260483P.
XX
PA   (BAND/) BANDMAN O.
XX
PI   Bandman O;
XX
DR   WPI; 2003-110597/10.
XX
PT   Combination for diagnosing, staging, treating, or monitoring the
PT   progression of treatment of a vascular disease, e.g. atherosclerosis,
PT   comprises several cDNAs that are differentially expressed in activated
PT   vascular tissue.
XX
PS   Claim 1; Page; 18pp; English.
XX
CC   This invention relates to a combination comprising several cDNAs that are
CC   differentially expressed in activated vascular tissue. The invention also
CC   discloses a high throughput method for detecting differentially expressed
CC   cDNAs in a sample. The cDNAs of the invention may have
CC   antiarteriosclerotic; cytostatic; cardiant; hypotensive; antidiabetic;
CC   gynaecological; vasotropic and cerebroprotective activities and may be
CC   used in gene therapy. The cDNAs of the invention may be used in a high-
CC   throughput methods for detecting differential expression of one or more
CC   cDNAs in a sample, or screening several molecules or compounds to
CC   identify a molecule or compound that specifically binds a cDNA of the
CC   invention. A protein encoded by the cDNA may be used to screen several
CC   molecules or compounds to identify a ligand that specifically binds to
CC   the protein, or to produce or purify an antibody to the protein that can
CC   be used to detect a protein in a sample or purify a natural or
CC   recombinant protein from a sample. The nucleotides may be useful for
CC   diagnosing, staging, treating, or monitoring the progression of treatment
CC   of a vascular disease, e.g. atherosclerosis, cancer, coronary artery
CC   disease, hypertension, diabetes, pre-eclampsia, ischaemia-reperfusion
CC   injury, restenosis, or stroke. The cDNAs can also be used for large-scale
CC   genetic or gene expression analysis of several new nucleic acid
CC   molecules. Antibodies to the proteins encoded by the cDNAs are useful for
CC   diagnosing pre-pathologic disorders, and chronic or acute diseases

```
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 TACCACTGTTGGTATAGGACTAAACTGACCAGATTCATGAAACAAAGCGCCATGAAG    480

Qy   497 TTTTAACTAGGACTGCTTCCTCGGCTGAGGATTCAATCCAGAGAAAATTAATTTGGACT  556
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TTTTAACTAGGACTGCTTCCTCGGCTGAGGATTCAATCCAGAGAAAATTAATTTGGACT  540

Qy   557 GCACAAATCCACGGTACTTAGCTGCGAGGTTCTTCCAATGCTGTCAAGCTCAGTAGGTTTT 616
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GCACAAATCCACGGTACTTAGCTGCGAGGTTCTTCCAATGCTGTCAAGCTCAGTAGGTTTT 600

Qy   617 TTGATAAGGTCATAGAGAAGAGAGATACTTTTTAAGAGAAGAAGTAATCAGATTCCAGGCTTTC 676
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TTGATAAGGTCATAGAGAAGAGAGATACTTTTTAAGAGAAGAAGTAATCAGATTCCAGGCTTTC 660

Qy   677 CTGATCGGAAAAGATTGGGGCAGGAACATTAAAGAGCCTTCATTATGACTGAATTACAC  736
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CTGATCGGAAAAGATTGGGGCAGGAACATTAAAGAGCCTTCATTATGACTGAATTACAC  720

Qy   737 TCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTCACAGCTCATCTTGCTGTGTTC  796
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 TCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTCACAGCTCATCTTGCTGTGTTC  780

Qy   797 AATTATTACCATCACAGCCTTTAACAAATCATCTAAAATGTCACCCTTCAGCCTTAC    856
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 AATTATTACCATCACAGCCTTTAACAAATCATCTAAAATGTCACCCTTCAGCCTTAC    840

Qy   857 CCTTTAATGAAAAATGAAAGGAAGTGACAATACGGAGGTCCAAACTTTGTCCCTGTTC  916
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 CCTTTAATGAAAAATGAAAGGAAGTGACAATACGGAGGTCCAAACTTTGTCCCTGTTC  900

Qy   917 TCTGTGTTCCTTACCTTCTGTCCCTGTGTATAGATTATGTAAAAGCCTTGTGTAAATAT  976
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 TCTGTGTTCCTTACCTTCTGTCCCTGTGTATAGATTATGTAAAAGCCTTGTGTAAATAT  960

Qy   977 GAGATGTGTGTCAAAATGATCAGTAGTAAATGAGCAATGACAGTGACTGCAGAGAAATTTA 1036
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 GAGATGTGTGTCAAAATGATCAGTAGTAAATGAGCAATGACAGTGACTGCAGAGAAATTTA 1020

Qy  1037 CTCTTGCCTAGAACTGGAGGGTTTTTATGGGTCTCGTAATTTTCCCACACTCATTGCTGAA 1096
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 CTCTTGCCTAGAACTGGAGGGTTTTTATGGGTCTCGTAATTTTCCCACACTCATTGCTGAA 1080

Qy  1097 AGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTACCTTCTTGAGGGGAACGGT 1156
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 AGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTACCTTCTTGAGGGGAACGGT 1140

Qy  1157 CTTGTTAACCAGCCCTGAGTTGTCTCACCCCAAACAATCTCTGTCATTTTCAAAGATGCAA 1216
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 CTTGTTAACCAGCCCTGAGTTGTCTCACCCCAAACAATCTCTGTCATTTTCAAAGATGCAA 1200

Qy  1217 AATGGTGTTATTTAATTGTCTCCACCATTGTCACACACAGGAATGCCTAATAATAGCAAC 1276
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 AATGGTGTTATTTAATTGTCTCCACCATTGTCACACACAGGAATGCCTAATAATAGCAAC 1260

Qy  1277 CCTTGTCTCCCCTCTTGCAAATGGCTCAGTGACTGGAAGAGGCGGACTAATA        1336
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 CCTTGTCTCCCCTCTTGCAAATGGGTCAGTGACTGGAAGAGGCGGACTAATA        1320

Qy  1337 GCCAGAGTAAAATACAAATTAATAATACATAGAGAACAGCAATACCAGAAAAAA      1396
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 GCCAGAGTAAAATACAAATTAATAATACATAGAGAACAGCAATACCAGAAAAAA      1380

Qy  1397 AGAATTCTGGTAAAATGATGTGAAAAATTGACAGCTCCCTCACTCTTAAGGTTGCTGCTA 1456
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 AGAATTCTGGTAAAATGATGTGAAAAATTGACAGCTCCCTCACTCTTAAGGTTGCTGCTA 1440

Qy  1457 TATACAGTCTAGGTTTTCTGTTTGAAAATAGGTAGGGTAAAATCTAAGACCTGCACAAGG 1516
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 TATACAGTCTAGGTTTTCTGTTTGAAAATAGGTAGGGTAAAATCTAAGACCTGCACAAGG 1500

Qy  1517 GCAGTGAGAGACATTTACAGCCTCCTCTCTCATTTGTTTTTTTAAGGAAAAGTCAACTCCT 1576
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

CC associated with abnormalities in the expression, amount or distribution
CC of the protein. The present sequence represents a cDNA of the invention
CC that is differentially expressed in activated vascular tissue. Note: The
CC sequence data for this patent did not form part of the specification, but
CC was obtained in electronic format directly from USPTO at
CC http.seqdata.uspto.gov/sequence.html?DocID=20020137081
XX
SQ Sequence 1417 BP; 426 A; 279 C; 308 G; 404 T; 0 U; 0 Other;

Query Match   80.2%; Score 1358.8; DB 8; Length 1417;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1360; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy   17 AGATGGGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAAT 76
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AGATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAAT 60

Qy   77 CTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCATCATGGACTGTGG 136
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCATCATGGACTGTGG 120

Qy   137 GCTGCTCAGTGAGTGTCAGAGTTCATGATTACAAAGAAGAACCCCAGAAGAAGAGACCT 196
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GCTGCTCAGTGAGTGTCAGAGTTCATGATTACAAAGAAGAACCCCAGAAGAAGAGACCT 180

Qy   197 ACTACTAGAATTATGGGATGTTGGAGGCTCGTGGGCAGTGCCAGCAGCGTGAAAAGCA 256
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ACTACTAGAATTATGGGATGTTGGAGGCTCGTGGGCAGTGCCAGCAGCGTGAAAAGCA 240

Qy   257 CAAGAGCAGTATTCTACAACTCCGTAAAATGTATTATTTTCGTACACGACTTAACAAATA 316
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CAAGAGCAGTATTCTACAACTCCGTAAAATGTATTATTTTCGTACACGACTTAACAAATA 300

Qy   317 AGAAGTCCTCCCAAAACTTGCCTCGTTGGTCGTTGGTCATTGGAAGCTCTCAACAGGGGATTTTGGTGC 376
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AGAAGTCCTCCCAAAACTTGCCTCGTTGGTCGTTGGTCATTGGAAGCTCTCAACAGGGGATTTTGGTGC 360

Qy   377 CAACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTGCTGATAACCAAA 436
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CAACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTGCTGATAACCAAA 420

Qy   437 TACCACTGTTGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAG 496
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 TACCACTGTTGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAG 480

Qy   497 TTTTAACTAGGACTGCTTCCTGGCGATTCAATCCAGAAGAAATTAATTTGGACT 556
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TTTTAACTAGGACTGCTTCCTGGCGATTCAATCCAGAAGAAATTAATTTGGACT 540

Qy   557 GCACAAATCCACGGTACTTAGCTGCAGTTCTTCCCAATGCGTCAAGCTCAGTAGGTTTT 616
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GCACAAATCCACGGTACTTAGCTGCAGTTCTTCCCAATGCGTCAAGCTCAGTAGGTTTT 600

Qy   617 TTGATAAGGTCATAGAAGAGATACTTTTTAAGAAGGTAATCAGATTCCAGGCCTTC 676
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TTGATAAGGTCATAGAAGAGATACTTTTTAAGAAGGTAATCAGATTCCAGGCCTTC 660

Qy   677 CTGATCGGAAAAGATTTGGGCGGCCAGGAACATTAAAGAGCCTTCATTATGACTGAATTACAC 736
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 CTGATCGGAAAAGATTTGGGCGGCCAGGAACATTAAAGAGCCTTCATTATGACTGAATTACAC 720

Qy   737 TCATCCTTGGAAGAGTGAGCAAGCAGCAAGCAGTGGCAGTTTTTCACAGCTCATCTTGCTGTGTTC 796
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 TCATCCTTGGAAGAGTGAGCAAGCAGCAAGCAGTGGCAGTTTTTCACAGCTCATCTTGCTGTGTTC 780

Qy   797 AATTATTACCATCACAGCCTTTAACAAAATCATCTTAAAATGCTACCCTTCAGCCTTAC 856
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 AATTATTACCATCACAGCCTTTAACAAAATCATCTTAAAATGCTACCCTTCAGCCTTAC 840

Qy   857 CCTTTAATGAAAAATGAAAGGAAGGAGTTGACAATACGGAGGTCCAAACTTTGTCCCTGTTC 916
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 CCTTTAATGAAAAATGAAAGGAAGGAGTTGACAATACGGAGGTCCAAACTTTGTCCCTGTTC 900

Qy   917 TCTGTGTTCCTTACCTTTCTGTCCCTGTGTATAGATTATGTAAAAAGCCTTGTGTAAAATAT 976
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 TCTGTGTTCCTTACCTTTCTGTCCCTGTGTATAGATTATGTAAAAAGCCTTGTGTAAAATAT 960

Qy   977 GAGATGTTGTCAAAATGATGCAGTAAATGAGCAATGACAGTGTACTGACAGAGAAAAATTTA 1036
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 GAGATGTTGTCAAAATGATGCAGTAAATGAGCAATGACAGTGTACTGACAGAGAAAAATTTA 1020

Qy   1037 CTCTTGCCTAGAACTGGAGGGTTTTATGGGTCGTGTAATTTCCCACACTCATTGCTGAA 1096
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CTCTTGCCTAGAACTGGAGGGTTTTATGGGTCGTGTAATTTCCCACACTCATTGCTGAA 1080

Qy   1097 AGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTACCTTCTTGAGGGGAACGGT 1156
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 AGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTACCTTCTTGAGGGGAACGGT 1140

Qy   1157 CTTGTTAACCAGCCCGAGTTGTCTACCCCAAACAATCTCCGTCATTTTCAAAGATGCAA 1216
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CTTGTTAACCAGCCCGAGTTGTCTACCCCATACAATCTCTCGTCATTTTCAAAGATGCAA 1200

Qy   1217 AATGGTGTTATTTAATTGTCTCCACCATTGTCACACAGGAAATGCCTAATAATAGCAAC 1276
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 AATGGTGTTATTGAATTGTCTCCACCATTGTCACACAGGAAATGCCTAATAATAGCAAC 1260

Qy   1277 CCTTGTCTCCCTCTTCTCTCCTTTGCAAATGGCTCAGTGACTGGAAGAGGCGGACTAATA 1336
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CCTTGTCTCCCTCTTCTCTCCTTTGCAAATGGCTCAGTGACTGGAAGAGGCGGACTAATA 1320

Qy   1337 GCCAGAGTTAAATATAAAATACAAATTAATAATAATACATAGAGAA 1378
           ||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 GCCAGAGTTAAATATAAAATACAAATTAATAATAATACATAGAGAA 1362

RESULT 4
ADE47763
ID   ADE47763 standard; DNA; 735 BP.
XX
AC   ADE47763;
XX
DT   29-JAN-2004 (first entry)
XX
DE   Human NOV38a gene SEQ ID NO:125.
XX
KW   ds; gene; human; cardiant; antiarteriosclerotic; hypotensive;
KW   immunosuppressive; dermatological; anorectic; cytostatic; antidiabetic;
KW   haemostatic; anti-HIV; antiasthmatic; antibacterial; virucide;
KW   neuroprotective; nootropic; antiparkinsonian; antilipaemic; gene therapy;
KW   vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200307642-A2.
XX
PD   18-SEP-2003.
XX
PF   02-AUG-2002; 2002WO-US024459.
XX
PR   02-AUG-2001; 2001US-0309501P.
PR   03-AUG-2001; 2001US-0310291P.
PR   08-AUG-2001; 2001US-0310951P.
PR   09-AUG-2001; 2001US-0311292P.
PR   13-AUG-2001; 2001US-0311979P.
PR   14-AUG-2001; 2001US-0312203P.
PR   17-AUG-2001; 2001US-0313156P.
PR   17-AUG-2001; 2001US-0313201P.
PR   20-AUG-2001; 2001US-0313702P.
PR   21-AUG-2001; 2001US-0314031P.
PR   23-AUG-2001; 2001US-0314466P.
PR   28-AUG-2001; 2001US-0315403P.
PR   29-AUG-2001; 2001US-0315853P.
PR   31-AUG-2001; 2001US-0316508P.
PR   21-SEP-2001; 2001US-0323936P.
PR   03-DEC-2001; 2001US-0338078P.

```
PR   05-FEB-2002; 2002US-0354655P.
PR   05-MAR-2002; 2002US-0361764P.
PR   19-APR-2002; 2002US-0373825P.
PR   15-MAY-2002; 2002US-0380971P.
PR   15-MAY-2002; 2002US-0380980P.
PR   16-MAY-2002; 2002US-0381039P.
PR   28-MAY-2002; 2002US-0383761P.
PR   29-MAY-2002; 2002US-0383887P.
PR   01-AUG-2002; 2002US-00210130.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Zerhusen BD, Patturajan M, Kekuda R, Miller CE, Rieger DK;
PI   Pena CEA, Shimkets RA, Li L, Berghs C, Zhong M, Casman SJ, Voss EZ;
PI   Boldog FL, Padigaru M, Smithson G, Shenoy SG, Ji W, Gorman L;
PI   Vernet CAM, Leite MW, Guo X, Anderson DW, Spytek KA, Gerlach VL;
PI   Burgess CE, Khrantsov NV, Ort T, Ellerman K, Rastelli L, Agee ML;
PI   Chaudhuri A, Chant JS, Dipippo VA, Edinger SR, Eisen A, Gangolli EA;
PI   Giot L, Ooi CE, Rothenberg ME, Spaderna SK, Hjalt T, Liu X;
PI   Taupier RJ, Catterton E;
XX
DR   WPI; 2003-779062/73.
DR   P-PSDB; ADE47764.
XX
PT   New NOVX polypeptides and nucleic acids, useful for preventing or
PT   treating NOVX-associated disorders, e.g. cancer, diabetes,
PT   atherosclerosis, asthma or AIDS, and in chromosome mapping, tissue typing
PT   or pharmacogenomics.
XX
PS   Claim 20; SEQ ID NO 125; 562pp; English.
XX
CC   The invention relates to a novel (NOVX) human polypeptide. A polypeptide
CC   of the invention has cardiant, antiarteriosclerotic, hypotensive,
CC   immunosuppressive, dermatological, anorectic, cytostatic, antidiabetic,
CC   haemostatic, anti-HIV, antiasthmatic, antibacterial, virucide,
CC   neuroprotective, nootropic, antiparkinsonian, and antilipaemic activity.
CC   A polynucleotide encoding a polypeptide of the invention may have a use
CC   in gene therapy, and as a vaccine. A polypeptide of the invention is
CC   useful in the manufacture of a medicament for treating a syndrome
CC   associated with a human disease, the disease selected from a pathology
CC   associated with the polypeptide. These may also be used in diagnosing,
CC   treating or preventing NOVX-associated disorders such as cardiomyopathy,
CC   atherosclerosis, hypertension, scleroderma, obesity, cancer, diabetes,
CC   haemophilia, graft-versus-host disease, AIDS, asthma, Crohn's disease,
CC   multiple sclerosis, infections, anorexia, cancer-associated cachexia,
CC   neurodegenerative disorders (e.g. Alzheimer's disease or Parkinson's
CC   disease), haematopoietic disorders, dyslipidaemias and other wasting
CC   disorders associated with chronic diseases. The nucleic acids are also
CC   used as hybridisation probes, in chromosome mapping, tissue typing,
CC   preventive medicine, and pharmacogenomics. The polypeptides are also
CC   useful as vaccines. The present sequence encodes a NOVX polypeptide of
CC   the invention.
XX
SQ   Sequence 735 BP; 220 A; 138 C; 183 G; 194 T; 0 U; 0 Other;

Query Match             43.4%;  Score 735;  DB 10;  Length 735;
Best Local Similarity   100.0%;  Pred. No. 1.2e-188;
Matches  735;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        10 AGTGGCAAGATGGGCGTCCCTGGAATCGGGTGAAGGTACTGGGTGTTGGGAGACTCAGGTGTT 69
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 AGTGGCAAGATGGGCGTCCCTGGAATCGGGTGAAGGTACTGGGTGTTGGGAGACTCAGGTGTT 60

Qy        70 GGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTCGGAAATCCATCATGG 129
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 GGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTCGGAAATCCATCATGG 120

Qy       130 ACTGTGGGCGTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAGGACCCCAGAAGAG 189
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 ACTGTGGGCGTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAGGACCCCAGAAGAG 180

Qy       190 AAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTG 249
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 AAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTG 240

Qy       250 AAAGCACCAAGAGCAGTATTCTACAACTCCGTAAATGTGTTATTTCGTACACGACTTA 309
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 AAAGCACCAAGAGCAGTATTCTACAACTCCGTAAATGTGTTATTTCGTACACGACTTA 300

Qy       310 ACAAATAAGAGAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGAT 369
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 ACAAATAAGAGAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGAT 360

Qy       370 TTGGTGCCAACTGGACTCTTGGTGCAAATGGGATTATGATCAAGAACAGTTTGCTGAT 429
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 TTGGTGCCAACTGGACTCTTGGTGCAAATGGGATTATGATCAAGAACAGTTTGCTGAT 420

Qy       430 AACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGC 489
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 AACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGC 480

Qy       490 CATGAAGTTTAACTAGGAGCTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAAT 549
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 CATGAAGTTTAACTAGGAGCTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAAT 540

Qy       550 TTGGACTGCACAAATCCACGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCAGT 609
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 TTGGACTGCACAAATCCACGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCAGT 600

Qy       610 AGGTTTTTGATAAGGTCATAGAGAGAGAGATACTTTTTAAGAGAAGGTAATCAGATTCCA 669
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 AGGTTTTTGATAAGGTCATAGAGAGAGAGATACTTTTTAAGAGAAGGTAATCAGATTCCA 660

Qy       670 GGCTTTCCTGATCGGAAAAGATTTGGGCGCAGAACATTAAAGAGCCTTCATTATGACTGA 729
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 GGCTTTCCTGATCGGAAAAGATTTGGGCGCAGAACATTAAAGAGCCTTCATTATGACTGA 720

Qy       730 ATTACACTCATCCTT 744
             |||||||||||||||
Db       721 ATTACACTCATCCTT 735

RESULT 5
ADJ79033
ID   ADJ79033 standard; DNA; 735 BP.
XX
AC   ADJ79033;
XX
DT   06-MAY-2004 (first entry)
XX
DE   Human NOVX protein Nov38A gene sequence.
XX
KW   NOVX; cytostatic; antidiabetic; anorectic; cerebroprotective;
KW   neuroprotective; antiinflammatory; thyromimetic; cardiant; gene-therapy;
KW   antisense-therapy; cancer; diabetes; obesity; endocrine disorder;
KW   CNS disorder; cardiovascular disorder; inflammatory disorder;
KW   detection assay; screening assay; chromosome mapping; tissue typing;
KW   predictive medicine; human; Nov38A; gene; ds.
XX
OS   Homo sapiens.
XX
PN   US2004014053-A1.
XX
PD   22-JAN-2004.
XX
PF   01-AUG-2002; 2002US-00210130.
XX
PR   02-AUG-2001; 2001US-0309501P.
PR   03-AUG-2001; 2001US-0310291P.
PR   08-AUG-2001; 2001US-0310951P.
PR   09-AUG-2001; 2001US-0311292P.
PR   13-AUG-2001; 2001US-0311979P.
PR   14-AUG-2001; 2001US-0312203P.
PR   17-AUG-2001; 2001US-0313156P.
PR   17-AUG-2001; 2001US-0313201P.
PR   20-AUG-2001; 2001US-0313643P.
```

PR 20-AUG-2001; 2001US-0313702P.
PR 21-AUG-2001; 2001US-0314031P.
PR 23-AUG-2001; 2001US-0314466P.
PR 28-AUG-2001; 2001US-0315403P.
PR 29-AUG-2001; 2001US-0315853P.
PR 31-AUG-2001; 2001US-0316508P.
PR 17-SEP-2001; 2001US-0322716P.
PR 21-SEP-2001; 2001US-0323336P.
PR 03-DEC-2001; 2001US-0338078P.
PR 05-FEB-2002; 2002US-0354655P.
PR 05-MAR-2002; 2002US-0361764P.
PR 19-APR-2002; 2002US-0373825P.
PR 15-MAY-2002; 2002US-0380971P.
PR 15-MAY-2002; 2002US-0380980P.
PR 16-MAY-2002; 2002US-0381039P.
PR 28-MAY-2002; 2002US-0383761P.
PR 29-MAY-2002; 2002US-0383887P.
XX
PA (ZERH/) ZERHUSEN B D.
PA (PATT/) PATTURAJAN M.
PA (KEKU/) KEKUDA R.
PA (MILL/) MILLER C E.
PA (RIEG/) RIEGER D K.
PA (PENA/) PENA C E A.
PA (SHIM/) SHIMKETS R A.
PA (LILL/) LI L.
PA (BERG/) BERGHS C.
PA (ZHON/) ZHONG M.
PA (CASM/) CASMAN S J.
PA (VOSS/) VOSS E Z.
PA (BOLD/) BOLDOG F L.
PA (PADI/) PADIGARU M.
PA (SMIT/) SMITHSON G.
PA (JIWW/) JI W.
PA (GORM/) GORMAN L.
PA (VERN/) VERNET C A M.
PA (LEIT/) LEITE M W.
PA (GUOX/) GUO X S.
PA (ANDE/) ANDERSON D W.
PA (SPYT/) SPYTEK K A.
PA (GERL/) GERLACH V.
PA (BURG/) BURGESS C E.
PA (KHRA/) KHRAMTSOV N V.
PA (ORTT/) ORT T.
PA (ELLE/) ELLERMAN K.
PA (RAST/) RASTELLI L.
PA (AGEE/) AGEE M L.
PA (CHAU/) CHAUDHURI A.
PA (CHAN/) CHANT J S.
PA (DIPI/) DIPIPPO V A.
PA (EDIN/) EDINGER S R.
PA (EISE/) EISEN A J.
PA (GANG/) GANGOLLI E A.
PA (GIOT/) GIOT L.
PA (OOIC/) OOI C E.
PA (ROTH/) ROTHENBERG M E.
PA (SPAD/) SPADERNA S K.
PA (HJAL/) HJALT T.
PA (LIUX/) LIU X.
PA (TAUP/) TAUPIER R J.
PA (CATT/) CATTERTON E.
PA (SHEN/) SHENOY S G.
XX
PI Zerhusen BD, Patturajan M, Kekuda R, Miller CE, Rieger DK;
PI Pena CEA, Shimkets RA, Li L, Berghs C, Zhong M, Casman SJ, Voss EZ;
PI Boldog FL, Padigaru M, Smithson G, Ji W, Gorman L, Vernet CAM;
PI Leite MW, Guo XS, Anderson DW, Spytek KA, Gerlach V, Burgess CE;
PI Khrantsov NV, Ort T, Ellerman K, Rastelli L, Agee ML, Chaudhuri A;
PI Chant JS, Dipippo VA, Edinger SR, Eisen AJ, Gangolli EA, Giot L;
PI Ooi CE, Rothenberg ME, Spaderna SK, Hjalt T, Liu X, Taupier RJ;
PI Catterton E, Shenoy SG;
XX
DR WPI; 2004-108206/11.

DR P-PSDB; ADJ79034.
XX
PT New isolated NOVX polypeptides and nucleic acid molecules useful for
PT treating, preventing and diagnosing pathological conditions with NOVX-
PT associated disorders, such as cancer, obesity, diabetes and inflammatory
PT or CNS diseases.
XX
PS Claim 20; SEQ ID NO 125; 250pp; English.
XX
CC This invention relates to a novel isolated NOVX polypeptide comprising a
CC fully defined sequence of, a mature form, one or more conservative
CC substitutions or at least 95% identity to 247 amino acids as given in the
CC specification. The invention may be useful for the development of
CC compounds with a cytostatic, antidiabetic, anorectic, cerebroprotective,
CC neuroprotective, antiinflammatory, thyromimetic or cardiant activity. In
CC addition, the disclosed sequences may prove useful for gene-therapy or
CC antisense-therapy. The invention may be useful for the diagnosis and
CC treatment of disorders associated with aberrant expression or activity of
CC the NOVX polypeptide, such as cancer, diabetes, obesity, and endocrine,
CC CNS, cardiovascular and inflammatory disorders. They can also be used in
CC various detection and screening assays, chromosome mapping, tissue typing
CC and predictive medicine. The present sequence is that of a gene which
CC encodes a human NOVX protein of the invention.
XX
SQ Sequence 735 BP; 220 A; 138 C; 183 G; 194 T; 0 U; 0 Other;

Query Match   43.4%; Score 735; DB 12; Length 735;
Best Local Similarity   100.0%; Pred. No. 1.2e-188;
Matches   735; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy   10 AGTGGCAAGATGGCGTCCCTGGATCGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTT 69
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 AGTGGCAAGATGGCGTCCCTGGATCGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTT 60

Qy   70 GGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATG 129
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATG 120

Qy  130 ACTGTGGGCTGCTCAGTGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAG 189
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 ACTGTGGGCTGCTCAGTGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAG 180

Qy  190 AAGACTTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTG 249
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 AAGACTTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTG 240

Qy  250 AAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGTATTATTTCGTACACGACTTA 309
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 AAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGTATTATTTCGTACACGACTTA 300

Qy  310 ACAAATAGAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGAT 369
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 ACAAATAGAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGAT 360

Qy  370 TTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTGCTGAT 429
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTGCTGAT 420

Qy  430 AACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGC 489
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 AACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGC 480

Qy  490 CATGGCAAGATGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAATTAAT 549
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 CATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAGAAATTAAT 540

Qy  550 TTGGACTGCACAAATCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCAGT 609
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 TTGGACTGCACAAATCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCAGT 600

Qy  610 AGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTTAAGAAGGTAATCAGATTCCA 669
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 AGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTTAAGAAGGTAATCAGATTCCA 660
```

```
Qy    670  GGCTTCCTGATCGGAAAAGATTTGGGCAGGAACATTATTAAGAGCCTTCATTATGACTGA 729
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  GGCTTCCTGATCGGAAAAGATTTGGGCAGGAACATTATTAAGAGCCTTCATTATGACTGA 720

Qy    730  ATTACACTCATCCTT 744
               |||||||||||||||
Db    721  ATTACACTCATCCTT 735
```

RESULT 6
ADE47765
ID   ADE47765 standard; DNA; 739 BP.
XX
AC   ADE47765;
XX
DT   29-JAN-2004 (first entry)
XX
DE   Human NOV38b gene SEQ ID NO:127.
XX
KW   ds; gene; human; cardiant; antiarteriosclerotic; hypotensive;
KW   immunosuppressive; dermatological; anorectic; cytostatic; antidiabetic;
KW   haemostatic; anti-HIV; antiasthmatic; antibacterial; virucide;
KW   neuroprotective; nootropic; antiparkinsonian; antilipaemic; gene therapy;
KW   vaccine.
XX
OS   Homo sapiens.
XX
PN   WO2003076642-A2.
XX
PD   18-SEP-2003.
XX
PF   02-AUG-2002; 2002WO-US024459.
XX
PR   02-AUG-2001; 2001US-0309501P.
PR   03-AUG-2001; 2001US-0310291P.
PR   08-AUG-2001; 2001US-0310951P.
PR   09-AUG-2001; 2001US-0311292P.
PR   13-AUG-2001; 2001US-0311979P.
PR   14-AUG-2001; 2001US-0312203P.
PR   17-AUG-2001; 2001US-0313156P.
PR   17-AUG-2001; 2001US-0313201P.
PR   20-AUG-2001; 2001US-0313702P.
PR   21-AUG-2001; 2001US-0314031P.
PR   23-AUG-2001; 2001US-0314466P.
PR   28-AUG-2001; 2001US-0315403P.
PR   29-AUG-2001; 2001US-0315853P.
PR   31-AUG-2001; 2001US-0316508P.
PR   21-SEP-2001; 2001US-0323936P.
PR   03-DEC-2001; 2001US-0338078P.
PR   05-FEB-2002; 2002US-0354655P.
PR   05-MAR-2002; 2002US-0361764P.
PR   19-APR-2002; 2002US-0373825P.
PR   15-MAY-2002; 2002US-0380971P.
PR   16-MAY-2002; 2002US-0381039P.
PR   28-MAY-2002; 2002US-0383761P.
PR   29-MAY-2002; 2002US-0383887P.
PR   01-AUG-2002; 2002US-00210130.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Zerhusen BD, Patturajan M, Kekuda R, Miller CE, Rieger DK;
PI   Pena CEA, Shimkets RA, Li L, Berghs C, Zhong M, Casman SJ, Voss EZ;
PI   Boldog FL, Padigaru M, Smithson G, Shenoy SG, Ji W, Gorman L;
PI   Vernet CAM, Leite MW, Guo X, Anderson DW, Spytek KA, Gerlach VL;
PI   Burgess CE, Khramtsov NV, Ort T, Ellerman K, Rastelli L, Agee ML;
PI   Chaudhuri A, Chant JS, Dippipo VA, Edinger SR, Eisen A, Gangolli EA;
PI   Giot L, Ooi CE, Rothenberg ME, Spaderna SK, Hjalt T, Liu X;
PI   Taupier RJ, Catterton E;
XX
DR   WPI; 2003-779062/73.
DR   P-PSDB; ADE47766.
```

XX
PT    New NOVX polypeptides and nucleic acids, useful for preventing or
PT    treating NOVX-associated disorders, e.g. cancer, diabetes,
PT    immunosuppressive, dermatological, anorectic, cytostatic, antidiabetic,
PT    atherosclerosis, asthma or AIDS, and in chromosome mapping, tissue typing
PT    or pharmacogenomics.
XX
PS    Claim 20; SEQ ID NO 127; 562pp; English.
XX
CC    The invention relates to a novel (NOVX) human polypeptide. A polypeptide
CC    of the invention has cardiant, antiarteriosclerotic, hypotensive,
CC    immunosuppressive, dermatological, anorectic, cytostatic, antidiabetic,
CC    haemostatic, anti-HIV, antiasthmatic, antibacterial, virucide,
CC    neuroprotective, nootropic, antiparkinsonian, and antilipaemic activity.
CC    A polynucleotide encoding a polypeptide of the invention may have a use
CC    in gene therapy, and as a vaccine. A polypeptide of the invention is
CC    useful in the manufacture of a medicament for treating a syndrome
CC    associated with a human disease. the disease selected from a pathology
CC    associated with the polypeptide. These may also be used in diagnosing,
CC    treating or preventing NOVX-associated disorders such as cardiomyopathy,
CC    atherosclerosis, hypertension, scleroderma, obesity, cancer, diabetes,
CC    haemophilia, graft-versus-host disease, AIDS, asthma, Crohn's disease,
CC    multiple sclerosis, infections, anorexia, cancer-associated cachexia,
CC    neurodegenerative disorders (e.g. Alzheimer's disease or Parkinson's
CC    disease), haematopoietic disorders, dyslipidaemias and other wasting
CC    disorders associated with chronic diseases. The nucleic acids are also
CC    used as hybridisation probes, in chromosome mapping, tissue typing,
CC    preventive medicine, and pharmacogenomics. The polypeptides are also
CC    useful as vaccines. The present sequence encodes a NOVX polypeptide of
CC    the invention.
XX
SQ    Sequence 739 BP; 222 A; 140 C; 185 G; 192 T; 0 U; 0 Other;

Query Match            43.2%;  Score 731.6;  DB 10;  Length 739;
Best Local Similarity   99.5%;  Pred. No. 1e-187;
Matches  734;  Conservative  0;  Mismatches  4;  Indels  0;  Gaps  0;

```
Qy     10  AGTGGCAAGATGGCGTCCCTGATCGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTT 69
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  AGTGGCAAGATCGCGTCCCTGATCGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTT 60

Qy     70  GGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGG 129
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  GGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGG 120

Qy    130  ACTGTGGGCTGCTCAGTGGATGTTGGAGGCTCGTGTGGGCAGTGCCAGCAGCGTG 189
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  ACTGTGGGCTGCTCAGTGGATGTTGGAGGCTCGTGTGGGCAGTGCCAGCAGCGTG 180

Qy    190  AAGACCTACTACATAGAATTATGGGAGGTTGGGAGGCTCGTTGGTCATTGGAAGCTCTCAACAGGGAT 249
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  AAGACCTACTACATAGAATTATGGGAGGTTGGGAGGCTCGTTGGTCATTGGAAGCTCTCAACAGGGAT 240

Qy    250  AAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTA 309
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  AAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTA 300

Qy    310  ACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTTCGTTGGTCATTGGAAGCTCTCAACAGGGAT 369
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  ACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTTCGTTGGTCATTGGAAGCTCTCAACAGGGAT 360

Qy    370  TTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTGCTGAT 429
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  TTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTGCTGAT 420

Qy    430  AACCAAATACCACTGGTTGGTAATAGGGACTAAACTGACTAAACTGAAACAAAGCGC 489
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  AACCAAATACCACTGGTTGGTAATAGGGACTAAACTGACTAAACTGAAACAAAGCGC 480

Qy    490  CATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTCAATCCAGAAGAAATTAAT 549
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  CATGAAGTTTAACTAGGACTGCTTTTCCTGGCTGAGGATTCAATCCAGAAGAAATTAAT 540

Qy    550  TTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCTTCCAAATGCTGTCAAGCTCAGT 609
```

```
Db       541  ||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
              TTGGGACTGCACAAATCCACGGTACTACTTAGCTGCAGGTTCCTCCAATGCTGTCAAGCTCAGT  600

Qy       610  AGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAATCAGATTCCA  669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601  AGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAATCAGATTCCA  660

Qy       670  GGCTTTCCTGATCGGAAAAGATTTGGGCGCAGGAACATTAAAGAGCCCTTCATTATGACTGA  729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661  GGCTTTCCTGATCGGAAAAGATTTGGGGCGCAGGAACATTAAAGAGCCCTTCATTATGACTGA  720

Qy       730  ATTACACTCATCCTTTGG  747
              ||||||||||||||  ||
Db       721  ATTACACTCATCCTAAGG  738
```

RESULT 7
ADJ79035
ID   ADJ79035 standard; DNA; 739 BP.
XX
AC   ADJ79035;
XX
DT   06-MAY-2004 (first entry)
XX
DE   Human NOVX protein Nov38B gene sequence.
XX
KW   NOVX; cytostatic; antidiabetic; anorectic; cerebroprotective;
KW   neuroprotective; antiinflammatory; thyromimetic; cardiant; gene-therapy;
KW   antisense-therapy; cancer; diabetes; obesity; endocrine disorder;
KW   CNS disorder; cardiovascular disorder; inflammatory disorder;
KW   detection assay; screening assay; chromosome mapping; tissue typing;
KW   predictive medicine; human; Nov38B; gene; ds.
XX
OS   Homo sapiens.
XX
PN   US2004014053-A1.
XX
PD   22-JAN-2004.
XX
PF   01-AUG-2002; 2002US-00210130.
XX
PR   02-AUG-2001; 2001US-0309501P.
PR   03-AUG-2001; 2001US-0310291P.
PR   08-AUG-2001; 2001US-0310551P.
PR   09-AUG-2001; 2001US-0311292P.
PR   13-AUG-2001; 2001US-0311979P.
PR   14-AUG-2001; 2001US-0312203P.
PR   17-AUG-2001; 2001US-0313156P.
PR   17-AUG-2001; 2001US-0313201P.
PR   20-AUG-2001; 2001US-0313643P.
PR   20-AUG-2001; 2001US-0313702P.
PR   21-AUG-2001; 2001US-0314031P.
PR   23-AUG-2001; 2001US-0314466P.
PR   28-AUG-2001; 2001US-0315403P.
PR   29-AUG-2001; 2001US-0315853P.
PR   31-AUG-2001; 2001US-0316508P.
PR   17-SEP-2001; 2001US-0322716P.
PR   21-SEP-2001; 2001US-0323936P.
PR   03-DEC-2001; 2001US-0338078P.
PR   05-FEB-2002; 2002US-0354655P.
PR   05-MAR-2002; 2002US-0361764P.
PR   19-APR-2002; 2002US-0373825P.
PR   15-MAY-2002; 2002US-0380971P.
PR   15-MAY-2002; 2002US-0380980P.
PR   16-MAY-2002; 2002US-0381039P.
PR   28-MAY-2002; 2002US-0383761P.
PR   29-MAY-2002; 2002US-0383887P.
XX
PA   (ZERH/) ZERHUSEN B D.
PA   (PATT/) PATTURAJAN M.
PA   (KEKU/) KEKUDA R.
PA   (MILL/) MILLER C E.
PA   (RIEG/) RIEGER D K.
PA   (PENA/) PENA C E A.
PA   (SHIM/) SHIMKETS R A.
PA   (LILL/) LI L.
PA   (BERG/) BERGHS C.
PA   (ZHON/) ZHONG M.
PA   (CASM/) CASMAN S J.
PA   (VOSS/) VOSS E Z.
PA   (BOLD/) BOLDOG F L.
PA   (PADI/) PADIGARU M.
PA   (SMIT/) SMITHSON G.
PA   (JIWW/) JI W.
PA   (GORM/) GORMAN L.
PA   (VERN/) VERNET C A M.
PA   (LEIT/) LEITE M W.
PA   (GUOX/) GUO X S.
PA   (ANDE/) ANDERSON D W.
PA   (SPYT/) SPYTEK K A.
PA   (GERL/) GERLACH V.
PA   (BURG/) BURGESS C E.
PA   (KHRA/) KHRAMTSOV N V.
PA   (ORTT/) ORT T.
PA   (ELLE/) ELLERMAN K.
PA   (RAST/) RASTELLI L.
PA   (AGEE/) AGEE M L.
PA   (CHAU/) CHAUDHURI A.
PA   (CHAN/) CHANT J S.
PA   (DIPI/) DIPIPPO V A.
PA   (EDIN/) EDINGER S R.
PA   (EISE/) EISEN A J.
PA   (GANG/) GANGOLLI E A.
PA   (GIOT/) GIOT L.
PA   (OOIC/) OOI C E.
PA   (ROTH/) ROTHENBERG M E.
PA   (SPAD/) SPADERNA S K.
PA   (HJAL/) HJALT T.
PA   (LIUX/) LIU X.
PA   (TAUP/) TAUPIER R J.
PA   (CATT/) CATTERTON E.
PA   (SHEN/) SHENOY S G.
XX
PI   Zerhusen BD, Patturajan M, Kekuda R, Miller CE, Rieger DK;
PI   Pena CEA, Shimkets RA, Li L, Berghs C, Zhong M, Casman SJ, Voss EZ;
PI   Boldog FL, Padigaru M, Smithson G, Ji W, Gorman L, Vernet CAM;
PI   Leite MW, Guo XS, Anderson DW, Spytek KA, Gerlach V, Burgess CE;
PI   Khramtsov NV, Ort T, Ellerman K, Rastelli L, Agee ML, Chaudhuri A;
PI   Chant JS, Dipippo VA, Edinger SR, Eisen AJ, Gangolli EA, Giot L;
PI   Ooi CE, Rothenberg ME, Spaderna SK, Hjalt T, Liu X, Taupier RJ;
PI   Catterton E, Shenoy SG;
XX
DR   WPI; 2004-108206/11.
DR   P-PSDB; ADJ79036.
XX
PT   New isolated NOVX polypeptides and nucleic acid molecules useful for
PT   treating, preventing and diagnosing pathological conditions with NOVX-
PT   associated disorders, such as cancer, obesity, diabetes and inflammatory
PT   or CNS diseases.
XX
PS   Claim 20; SEQ ID NO 127; 250pp; English.
XX
CC   This invention relates to a novel isolated NOVX polypeptide comprising a
CC   fully defined sequence of, a mature form, one or more conservative
CC   substitutions or at least 95% identity to 247 amino acids as given in the
CC   specification. The invention may be useful for the development of
CC   compounds with a cytostatic, antidiabetic, anorectic, cerebroprotective,
CC   neuroprotective, antiinflammatory, thyromimetic or cardiant activity. In
CC   addition, the disclosed sequences may prove useful for gene-therapy or
CC   antisense-therapy. The invention may be useful for the diagnosis and
CC   treatment of disorders associated with aberrant expression or activity of
CC   the NOVX polypeptide, such as cancer, diabetes, obesity, and endocrine,
CC   CNS, cardiovascular and inflammatory disorders. They can also be used in
CC   various detection and screening assays, chromosome mapping, tissue typing
CC   and predictive medicine. The present sequence is that of a gene which
CC   encodes a human NOVX protein of the invention.

```
XX
SQ   Sequence 739 BP; 222 A; 140 C; 185 G; 192 T; 0 U; 0 Other;

Query Match             43.2%;    Score 731.6;    DB 12;    Length 739;
Best Local Similarity   99.5%;    Pred. No. 1e-187;
Matches 734;   Conservative    0;   Mismatches    4;   Indels    0;   Gaps    0;

Qy   10 AGTGGCAAGATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTT 69
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 AGTGGCAAGATCGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTT 60

Qy   70 GGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGCTGCTGGGAAATCCATCATGG 129
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGCTGCTGGGAAATCCATCATGG 120

Qy  130 ACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAG 189
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 ACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAG 180

Qy  190 AAGACCTACTACATAGAATTATGGAGGCTCGTGGGCAGTGCCAGCAGCGTG 249
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 AAGACCTACTACATAGAATTATGGAGGCTCGTGGGCAGTGCCAGCAGCGTG 240

Qy  250 AAAAGCACAAGAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTA 309
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 AAAAGCACAAGAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTA 300

Qy  310 ACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGAT 369
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 ACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGAT 360

Qy  370 TTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGAT 429
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGAT 420

Qy  430 AACCAAATACCACTGTTGGTAATAGGACTGCTTTCCTGCTCAAGATTCAATCCAGAAGAAATTAAT 489
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 AACCAAATACCACTGTTGGTAATAGGACTGCTTTCCTGCTCAAGATTCAATCCAGAAGAAATTAAT 480

Qy  490 CATGAAGTTTTAACTAGGACTGCTTTCCTGCTCAGGTTCTTCCAATGCTGTCAAGCTCAGT 549
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 CATGAAGTTTTAACTAGGACTGCTTTCCTGCTCAGGTTCTTCCAATGCTGTCAAGCTCAGT 540

Qy  550 TTGGACTGCACAAATCCACCGTACTTAGCTGCAGTTCTTCCCAATGCTGTCAAGCTCAGT 609
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 TTGGACTGCACAAATCCACCGTACTTAGCTGCAGTTCTTCCCAATGCTGTCAAGCTCAGT 600

Qy  610 AGTTTTTTGATAAGGTCATAGAGAAGAGATACTTTTAAGAGAAGGTAATCAGATTCAGATTCCA 669
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 AGTTTTTTGATAAGGTCATAGAGAAGAGATACTTTTAAGAGAAGGTAATCAGATTCAGATTCCA 660

Qy  670 GGCTTCCTGATCGGAAAAGATTTGGGGCAGGAAACATTAAAGAGCCTTCATTATGACTGA 729
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 GGCTTCCTGATCGGAAAAGATTTGGGGCAGCAAACATTAAAGAGCCTTCATTATGACTGA 720

Qy  730 ATTACACTCATCCTTTGG 747
         ||||||||||||||||
Db  721 ATTACACTCATCCTAAGG 738


RESULT 8
AAS27286
ID   AAS27286 standard; cDNA; 959 BP.
XX
AC   AAS27286;
XX
DT   07-NOV-2001  (first entry)
XX
DE   cDNA encoding novel signal transduction pathway protein, Seq ID 321.
XX
KW   Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW   antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW   immune system disorder; rheumatoid arthritis; inflammatory condition;
```

PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
PR   05-DEC-2000; 2000US-0251030P.
PR   05-DEC-2000; 2000US-0251988P.
PR   05-DEC-2000; 2000US-0256719P.
PR   06-DEC-2000; 2000US-0251479P.
PR   08-DEC-2000; 2000US-0251856P.

PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
PR   08-DEC-2000; 2000US-0251989P.
PR   08-DEC-2000; 2000US-0251990P.
PR   11-DEC-2000; 2000US-0254097P.
PR   05-JAN-2001; 2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-465460/50.
DR   P-PSDB; AAU17369.
XX
PT   Novel polypeptides useful for diagnosing, treating, preventing and/or
PT   prognosing disorders related to the proteins, including cancers, immune
PT   disorders and neuronal disorders.
XX
PS   Claim 1; SEQ ID NO 321; 880pp; English.
XX
CC   The invention relates to novel isolated polypeptides (I), and
CC   polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC   diagnosing, preventing and treating diseases including immune system
CC   disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC   disorders (e.g rheumatoid arthritis), inflammatory conditions, organ
CC   transplant rejections and graft versus host disease, infectious diseases
CC   (e.g. hepatitis C), bleeding disorders (sickle cell anaemia), haemoglobin abnormalities and
CC   other blood-related disorders (sickle cell anaemia), myeloproliferative
CC   disorders, primary haematopoietic disorders, hyperproliferative disorders
CC   (e.g. Gaucher's disease and cancer), neurodegenerative disorders (e.g.
CC   Alzheimer's disease, Parkinson's disease), chromosomal abnormalities
CC   (Down syndrome), ischaemic injury (e.g. stroke), renal disorders (e.g.
CC   glomerulonephritis), cardiovascular disorders (e.g. arrhythmia),
CC   respiratory disorders, dermatological disorders, in wound healing,
CC   epithelial cell proliferation, endocrine disorders (e.g. Addison's
CC   disease), reproductive system disorders, gastrointestinal disorder
CC   (inflammatory disorders), liver disorders (cirrhosis), as stimulators of
CC   B-cell responsiveness to pathogens, activators of T-cells, to induce
CC   higher affinity antibodies, and as a means to induce tumour proliferation
CC   in pathologies e.g. acquired immune deficiency syndrome (AIDS). AAS26976-
CC   AAS27850 represent novel signal transduction pathway protein coding
CC   sequences and PCR primers of the invention
XX

Query Match   37.1%; Score 628.4; DB 4; Length 959;
Best Local Similarity 96.5%; Pred. No. 1.1e-159;
Matches 684; Conservative   1; Mismatches 18; Indels 6; Gaps 4;

Qy   8 GCAGTGGCAAGATGGCCTCCCTGGATCGGGTGAAGGTACTCGGTGTGTTGGGAGACTCAGGTG 67
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GCAGTGGCAAGATGGGCGTCCCTGGATCGGGTGAAGGTACTCGGTGTGTTGGGAGACTCAGGTG 60

Qy   68 TTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCAT 127
            |||||||||||||||||||||||||||||||||||||||||
Db   61 TTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCAT 120

Qy   128 GGACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAAGACCCCAGAG 187
            ||||||||||||||||||||||||||||||||||||
Db   121 GGACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAAGACCCCAGAG 180

Qy   188 AGAAGACCTACTACATAGAATTATGGGATGTTGGAGGCTTCTGTGGGCCAGTGCCAGCAGCG 247
            ||||||||||||||||||||||||||||||||||||||
Db   181 AGAAGACCTACTACATAGAATTATGGGATGTTGGAGGCTTCTGTGGGCCAGTGCCAGCAGCG 240

Qy   248 TGAAAAGCACAAGAGACAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACT 307
            |||||||||||||||||||||||||||||||||||||||
Db   241 TGAAAAGCACAAGAGACAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACT 300

Qy   308 TAACAAATAAGAGAGTCCTCCCAAAACTTGCGTCGTTGGTCGTTGGTCATTGGAAGCTCTCAACAGGG 367
            |||||||||||||||||||||||||
Db   301 TAACAAATAAGAGAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGG 360

Qy   368 ATTTGGTGCCAACTGGAGTCTTCGTTGACAAATGGGGATTATGATCAAGAACAGTTTGCTG 427

```
Db      361  ATTTGGTGCCAACTGGAGTCTTGTGTGACAAATGGGATTATGATCAAGAACAGTTTGCTG  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy      428  ATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGC  487
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421  ATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGC  480
Qy      488  GCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTA  547
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481  GCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTA  540
Qy      548  ATTTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCA  607
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541  ATTTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAA-CTCA  599
Qy      608  GTAGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAATCAGATTC  667
             :||||||||||||||||  ||||||||||||||||||||||  ||||||||||||||||
Db      600  RTAGGTTTTTTTGAT-AGGGCATAGAGAGAAGAGATAC-TTTTAAGAGAAGGTAATCNGATCC  657
Qy      668  CAGGCTTCCTGATCGGAAAAAGATTTGGGCAGGAACATTAAAGAGCCT  716
             ||||||   ||  |||||||||||||||||||||||||||||||||||
Db      658  GAGGCTTCTTGACTCG---AAAAGATTGGGCCGGACCATTAAAGCGCTT  703
```

```
RESULT 9
ADB93464
ID   ADB93464 standard; cDNA; 959 BP.
XX
AC   ADB93464;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human cDNA encoding a novel protein #311.
XX
KW   ss; gene; human; autoimmune disease; Parkinson's disease; silicosis;
KW   gastrointestinal disease; atherosclerosis; adjuvant; thrombocytopenia;
KW   immunosuppressive agent; adjuvant; enhance immune response;
KW   higher affinity antibody induction;
KW   increased serum immunoglobulin concentration.
XX
OS   Homo sapiens.
XX
PN   US2002168711-A1.
XX
PD   14-NOV-2002.
XX
PF   17-JAN-2001; 2001US-00764868.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   28-JUN-2000; 2000US-0214886P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225454P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   22-AUG-2000; 2000US-0226868P.
PR   30-AUG-2000; 2000US-0289024P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   08-SEP-2000; 2000US-0211413P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   27-SEP-2000; 2000US-0235834P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241809P.
PR   01-NOV-2000; 2000US-0244617P.
PR   17-NOV-2000; 2000US-0249299P.
PR   08-DEC-2000; 2000US-0251856P.
PR   08-DEC-2000; 2000US-0251868P.
PR   08-DEC-2000; 2000US-0251869P.
XX
PA   (ROSE/) ROSEN C A.
PA   (RUBE/) RUBEN S M.
PA   (BARA/) BARASH S C.
XX
PI   Rosen CA, Ruben SM,  Barash SC;
XX
DR   WPI; 2003-719985/68.
DR   P-PSDB; ADB94077.
XX
PT   New isolated polypeptide useful for diagnosing and treating
PT   immunosuppressive conditions such as autoimmune disease and Parkinson's
PT   disease.
XX
PS   Claim 3; SEQ ID NO 321; 345pp; English.
XX
CC   The invention relates to an isolated polypeptide. The polypeptide is
CC   useful for diagnosing a pathological condition or a susceptibility to a
CC   pathological condition in a subject, by determining the presence or
CC   amount of expression of the polypeptide in a biological sample and
CC   diagnosing a pathological condition or a susceptibility to a pathological
CC   condition based on the presence or amount of expression of the
CC   polypeptide. The polypeptide is also useful for identifying a binding
CC   partner to the polypeptide, which involves contacting the polypeptide
CC   with a binding partner and determining whether the binding partner
CC   effects an activity of the polypeptide. The polypeptide or the nucleic
CC   acid encoding the polypeptide is useful for preventing, treating, or
CC   ameliorating a medical condition, which involves administering the
CC   polypeptide or the nucleic acid to a mammalian subject. The nucleic acid
CC   is useful for diagnosing a pathological condition or a susceptibility to
CC   a pathological condition in a subject, which involves determining the
CC   presence or absence of a mutation in the nucleic acid, and diagnosing a
CC   pathological condition or susceptibility to a pathological condition
CC   based on the presence or absence of the mutation. The polypeptide, the
CC   nucleic acid and an antibody to the polypeptide are useful for treating
CC   autoimmune disease, Parkinson's disease, silicosis, gastrointestinal
CC   disease, atherosclerosis, haemophilia, thrombocytopenia. The polypeptide,
CC   the nucleic acid and the antibody are useful as immunosuppressive agents,
CC   as adjuvants to enhance immune responses, and as agents to induce higher
CC   affinity antibodies and increase serum immunoglobulin concentrations. The
CC   present sequence represents cDNA encoding a novel human protein. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format direct from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20020168711.
XX
SQ   Sequence 959 BP; 260 A; 200 C; 249 G; 247 T; 0 U; 3 Other;

Query Match             37.1%;  Score 628.4;  DB 10;  Length 959;
```

Fri Jul 1 07:44:52 2005   us-09-945-173-1.rng

Best Local Similarity 96.5%; Pred. No. 1.1e-159;
Matches 684; Conservative 1; Mismatches 18; Indels 6; Gaps 4;

```
Qy    8 GCAGTGCCAAGATGGCGTCCCTGGATCGGGTGAAGGTACTGGTTGGGAGACTCAGGTG 67
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GCAGTGCCAAGATGGCGTCCCTGGATCGGGTGAAGGTACTGGTTGGGAGACTCAGGTG 60

Qy   68 TTGGGAAATCTTCGTTAGTCCATCTCCATGCCAAAATCAAGTGCTGGGAAATCCATCAT 127
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TTGGGAAATCTTCGTTAGTCCATCTCCATGCCAAAATCAAGTGCTGGGAAATCCATCAT 120

Qy  128 GGACTGTGGGCGTCGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAG 187
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GGACTGTGGGCGTCGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAG 180

Qy  188 AGAAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCG 247
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 AGAAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCG 240

Qy  248 TGAAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACT 307
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 TGAAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACT 300

Qy  308 TAACAAATAAGAGAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGG 367
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 TAACAAATAAGAGAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGG 360

Qy  368 ATTTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTG 427
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 ATTTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTG 420

Qy  428 ATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGC 487
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGC 480

Qy  488 GCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTA 547
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 GCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTA 540

Qy  548 ATTTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCA 607
          ||||||||||||||||||||||||||||| ||||| ||||||||||||||||||||||
Db  541 ATTTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAA-CTCA 599

Qy  608 GTAGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTAAGAGAAGGTAATCAGATTC 667
          :|||||||||||: |||| |||||||||||||||||: |||||| || ||||||||||
Db  600 RTAGGTTTTTTGAT-AGGGCATAGAGAAGAGAGATAC-TTTTAAGGAAGGTAATCNGATCC 657

Qy  668 CAGGGCTTCCTGATCGGAAAAGATTTGGGCAGGAACATTAAAGAGAGCT 716
          ||||||||||||||||| ||| | ||||| || ||||||||||||||||
Db  658 GAGGGCTTCTTGACTCG--AAAAGATTGGGCCGGACCATTAAAGAGCTT 703
```

RESULT 10
ACH36304
ID   ACH36304 standard; cDNA; 504 BP.
XX
AC   ACH36304;
XX
DT   13-OCT-2003 (first entry)
XX
DE   Human endothelial cell cDNA #4437.
XX
KW   Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW   genome mapping; biodiversity; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   US2003073623-A1.
XX
PD   17-APR-2003.
XX
PF   30-JUL-2001; 2001US-00918995.
XX

PT   New polynucleotide sequences obtained from various cDNA libraries, useful
PT   as hybridization probes, as oligomers for PCR, for chromosome and gene
PT   mapping, in the recombinant production of protein, or in generating
PT   antisense DNA or RNA.
XX
PS   Claim 1; SEQ ID NO 23516; 44pp; English.
XX
CC   The invention relates to an isolated polynucleotide comprising any one of
CC   38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC   determined by the technique of SBH (sequencing by hybridisation). Also
CC   included is a purified polypeptide comprising a sequence corresponding to
CC   a reading frame of the novel polynucleotide. The nucleic acid sequences
CC   are useful in diagnostics as expressed sequence tags (EST) for
CC   identifying expressed genes or for physical mapping of the human genome,
CC   in forensics, in assessing biodiversities, or in identifying mutations
CC   responsible for genetic disorders and other traits. The nucleotide
CC   sequences are also useful as hybridisation probes, as oligomers for PCR,
CC   for chromosome and gene mapping, in the recombinant production of
CC   protein, or in generating antisense DNA or RNA. The purified polypeptide
CC   is useful for generating antibodies specific for it. The present sequence
CC   is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=2003073623
XX
SQ   Sequence 504 BP; 148 A; 95 C; 132 G; 121 T; 0 U; 8 Other;

Query Match   26.2%; Score 444; DB 9; Length 504;
Best Local Similarity 96.4%; Pred. No. 8.6e-110;
Matches 450; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
Qy   20 TGGCGTCCCTGGATCGGGTGAAGGTACTGGTTGGGAGACTCAGGTGTTGGGAAATCTT 79
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   37 TTGGNANNNNGGCTCGGNNGAAGGGACTGGTTGGGAGACTCAGGTGTTGGGAAATCTT 96

Qy   80 CGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCAAGTGCTTAACAAATAAGA 139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   97 CGTTAGTCCATCTCCTAAATCAAGTGCTGGGAAATCAAGTGCTGGGAAATCAAGTGCTTAACAAATAAGA 156

Qy  140 GCTCAGTGGATGTCAGAGTTCATGATTACAAAGAACCCCAGAAGGAACCCCAGAAGACCTACT 199
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  157 GCTCAGTGGATGTCAGAGTTCATGATTACAAAGAACCCCAGAAGGAACCCCAGAAGACCTACT 216

Qy  200 ACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCCAGCAGCGTGAAAAGCACAA 259
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  217 ACATAATTATTATGGGATGTTGGAGGCTCTGTGGGCCAGCAGCGTGAAAAGCACAA 276

Qy  260 GAGCAGTATTCTACAACTCCGTAAATGTGTAAATGTGTAAATCAAGTTAACAAATAAGA 319
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  277 GAGCAGTATTCTACAACTCCGTAAATGTGTAAATCAAGTTAACAAATAAGA 336

Qy  320 AGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGCTGATAACCAA 379
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  337 AGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGCTGATAACCAA 396

Qy  380 CTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATAC 439
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  397 CTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATAC 456

Qy  440 CACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAG 486
          |||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    457 CACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAG 503

RESULT 11
AAX41283
ID   AAX41283 standard; cDNA; 385 BP.
XX
AC   AAX41283;
XX
DT   17-JUN-1999 (first entry)
XX
DE   Human secreted protein 5' EST SEQ ID NO:227.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS   Homo sapiens.
XX
PN   WO9906548-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998; 98WO-IB001222.
XX
PR   01-AUG-1997; 97US-00905135.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Lacroix B;
XX
DR   WPI; 1999-153778/13.
DR   P-PSDB; AAY12450.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from cDNA
PT   libraries prepared from e.g. liver, ovary, brain, prostate, kidney, lung,
PT   umbilical cord, placenta and colon tissue.
XX
PS   Claim 1; Page 509; 824pp; English.
XX
CC   AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY12261 to
CC   AAY12514, respectively. The proteins given represent the signal peptide
CC   and an N-terminal fragment of a secreted protein. The nucleic acid
CC   sequences can be used for producing secreted human gene products. They
CC   can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell
XX
SQ   Sequence 385 BP; 103 A; 74 C; 109 G; 91 T; 0 U; 8 Other;

Query Match             22.1%;  Score 374.8;  DB 2;  Length 385;
Best Local Similarity   97.4%;  Pred. No. 4.5e-91;
Matches  374;  Conservative   5;  Mismatches   5;  Indels   0;  Gaps   0;

Qy        10 AGTGGCAAGATGGCCGTCCCCTGGATCGGGTGAAGGTACTGGTGTTCGGAGACTCAGGTGTT 69
             ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db         2 AGTGGCAAGATGGCGCGTCCCCTGGATCGGGTGAAGGTACTGGTGTTCGGAGACTCAGGTGTT 61

Qy        70 GGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGG 129
             ||||||||||||||||||||||||||
```

```
Db        62 GGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGG 121
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy       130 ACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAACCCAGAAGAG 189
             |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       122 ACTGTGGGCTGCTCAGTGGATGTCAGAGTKCATGATTACAAAGAAGAACCCAGAAGAG 181

Qy       190 AAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCGTGTGGCCAGTGCCAGCGGTG 249
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       182 AAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCGTGTGGCCAGTGCCAGCGGTG 241

Qy       250 AAAAGACAACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTA 309
             ||||||||||||||||||||||||||||||||||||||||||||::||||
Db       242 AAAAGACAACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTWNYCGTACACGACTTA 301

Qy       310 ACAAATAAGAGAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGAT 369
             ||:||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       302 ACVSATGGGAAGTCCTCCCAAAAMTTGCGNCGTTGGTCATTGGAAGCTCTCAACAGGGAT 361

Qy       370 TTGGTGCCAACTGGAGTCTTGGTG 393
             ||||||||||||||||||||||||
Db       362 TTGGTGCCAACTGGAGTCTTGGTG 385

RESULT 12
AAC17946
ID   AAC17946 standard; cDNA; 321 BP.
XX
AC   AAC17946;
XX
DT   06-OCT-2000 (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 22021.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-00200610.
XX
PR   26-FEB-1999;  99US-0122487P.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS   Claim 1; SEQ ID NO 22021; 71pp + Sequence Listing; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC   ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC   to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC   diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors
XX
```

SQ Sequence 321 BP; 89 A; 62 C; 66 G; 104 T; 0 U; 0 Other;

Query Match   18.9%;   Score 321; DB 3; Length 321;
Best Local Similarity 100.0%;   Pred. No. 1.6e-76;
Matches 321;   Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   871 ATGAAAGGAAGTGACAATACGGGAGGTCCAAACTTTGTCCCGTTCTCTGTTCCTTAC 930
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAAAGGAAGTGACAATACGGGAGGTCCAAACTTTGTCCCGTTCTCTGTTCCTTAC 60

Qy   931 CTTTCTGTCCCGTGTATAGATTATGTAAAAGCCTTGTGTAAATATGAGATGTTGTCAAA 990
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CTTTCTGTCCCGTGTATAGATTATGTAAAAGCCTTGTGTAAATATGAGATGTTGTCAAA 120

Qy   991 ATGATGCAGTAAATGAGCAATGACACAGTGTACTGCAGAGAAAATTTACTCTTGCCTAGAAC 1050
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ATGATGCAGTAAATGAGCAATGACACAGTGTACTGCAGAGAAAATTTACTCTTGCCTAGAAC 180

Qy   1051 TGGAGGTTTTTATGGGTCTGTAATTTTCCCACACTCATTGCTGAAAGCTTAATTAAGTA 1110
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TGGAGGTTTTTATGGGTCTGTAATTTTCCCACACTCATTGCTGAAAGCTTAATTAAGTA 240

Qy   1111 CTTCAAAAACGTATCTCCATTGTTTACCTTCTTGAGGGAGGGACGGTCTTGTTAACCAGCC 1170
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CTTCAAAAACGTATCTCCATTGTTTACCTTCTTGAGGGAGGGACGGTCTTGTTAACCAGCC 300

Qy   1171 CTGAGTTGTCTACCCCAAACA 1191
          |||||||||||||||||||||
Db   301 CTGAGTTGTCTACCCCAAACA 321

RESULT 13
ABX52440
ID   ABX52440 standard; cDNA; 387 BP.
XX
AC   ABX52440;
XX
DT   25-FEB-2003 (first entry)
XX
DE   Bovine EST associated with lactation/muscle/fat deposition #2369.
XX
KW   Bovine; 88; EST; expressed sequence tag; lactation; LMFD;
KW   muscle deposition; fat deposition; genome mapping; gene identification;
KW   gene analysis; cattle breeding.
XX
OS   Bos Taurus.
XX
PN   US2002137160-A1.
XX
PD   26-SEP-2002.
XX
PF   26-OCT-2001; 2001US-00983965.
XX
PR   17-DEC-1998; 98US-0113678P.
PR   15-DEC-1999; 99US-00465231.
XX
PA   (BYAT/) BYATT J C.
PA   (MATH/) MATHIALAGAN N.
PA   (TAON/) TAO N.
PA   (WARR/) WARREN W C.
XX
PI   Byatt JC, Mathialagan N, Tao N, Warren WC;
XX
DR   WPI; 2003-102386/09.
XX
PT   Purified nucleic acid molecules, useful for genome mapping, gene
PT   identification and analysis, cattle breeding or preparation of constructs
PT   for cattle gene expression and genetically improved cattle.
XX
PS   Claim 2; SEQ ID NO 2369; 38pp; English.
XX
CC   The invention relates to a purified nucleic acid molecule associated with
CC   lactation or muscle and fat deposition (designated LMFD), derived from

CC   cattle, and the LMFD nucleic acid can specifically hybridise to a second
CC   nucleic acid molecule comprising any of 5912 nucleotide sequences,
CC   appearing as ABX50072-ABX55983, or complements of them. Also included are
CC   ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
CC   acid linked to a promoter and a 3', non-translated sequence that
CC   functions in the cell to cause termination of transcription and addition
CC   of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
CC   (2) determining a level or pattern of a molecule in a bovine cell or
CC   tissue comprising: (a) incubating a marker nucleic acid (comprising any
CC   of the 5912 nucleic acid sequences or its complement or fragment) with a
CC   complementary nucleic acid molecule obtained from the bovine cell or
CC   tissue, where hybridisation between the marker nucleic acid and the
CC   complementary nucleic acid permits the detection of the molecule; and (b)
CC   detecting the level or pattern of the complementary nucleic acid, where
CC   the detection of the complementary nucleic acid is predictive of the
CC   level or pattern of the molecule. The LMFD nucleic acid is used for
CC   determining a level or pattern of a molecule in a bovine cell or tissue.
CC   It is useful for genome mapping, gene identification and analysis, cattle
CC   breeding, preparation of constructs for use in cattle gene expression, or
CC   for genetically improving cattle. The present sequence is one of the 5912
CC   bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The present
CC   sequence was not shown in the specification but was obtained in
CC   electronic format from the USPTO web site:
CC   seqdata.uspto.gov/sequence.html?DocID=20020137160
XX
SQ   Sequence 387 BP; 107 A; 74 C; 75 G; 131 T; 0 U; 0 Other;

Query Match   14.9%;   Score 252.8; DB 8; Length 387;
Best Local Similarity 84.5%;   Pred. No. 5.4e-58;
Matches 333;   Conservative   0; Mismatches 52; Indels 9; Gaps   4;

Qy   760 GCAGTGGCAGTTTTCACAGCTCATCTTGCTGTGT-TCAATTATTACCATCACAGCCTTT 818
          | ||||||||||||| ||||||||||||||||||| ||||||| ||
Db   1 GCGGTGGCAGTTTTTCACAGCTTTCTTCTTGCATTAATGTCTCAGCCTTT 60

Qy   819 TAACAAAATCATC-TTAAAATGCTACCCTTCAGCCTTAAATGGAAAATGAAAG 877
          ||| ||||||| ||| |||||||| |||||||||||| ||||||||||||
Db   61 TAATGAAATCATCTTTAAAATGCCACCCTTCAGCCTCCCTTTAGTGGAAAACTGAAAG 120

Qy   878 GAAGTGACAATACGGGAGGTCCAAACTTTGTCCCGTTCTCTGTTCCTTACCTTTCTG 937
          |||||||||||||||||||||||||||||||||||||||||||| |||
Db   121 GAAGTAACACGGTGGGAGCTCCAGACTTGGTCC----TCATTCTATTTTTTCACCCTTTCTG 176

Qy   938 TCCCTGTGTATAGATTATTGTAAAAGCCTTGTGTAAAATGAGATGTTGTCAAAATGATGC 997
          ||||||||||||||||||||||||||||| ||||| |||||||||||||||
Db   177 TCCCTGTTTATAGATTATTATATAAAAACCTGTGGAAATATGA---TGTCAAAATGATGC 233

Qy   998 AGTAAATGAGCAATGACAGTGTACTGCAGAGAAAATTTACTCTTGCCTAGAACTGGAGGG 1057
          |||||||||||||||| || ||| |||||||||||||||||||||||
Db   234 AGTAAATGAGCAGTGACAGAATGCTGCAGAGAAAAATTTACTCTTGCCTAGAACTGGAGGA 293

Qy   1058 TTTTTATGGGTCTGTAATTTTCCCACACTCATTGCTGAAAGCTTAATTAAGTACTTCAAA 1117
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   294 TTTTTATGGGTCTGTAATTTTCCCACACTCATTGCTGAAAGCTTAATTAAGTACTTCAGA 353

Qy   1118 AACGTATCTCCATTGTTTACCTTCTTGAGGGA 1151
          || ||||||||||||||||||||||||||||
Db   354 AATGTATCCCATTGTTTTACCTTCTGAGGGGAA 387

RESULT 14
ABL08307
ID   ABL08307 standard; cDNA; 919 BP.
XX
AC   ABL08307;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 19403.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX

```
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB64204.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 19403; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB7737-
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB7737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 919 BP; 226 A; 258 C; 257 G; 178 T; 0 U; 0 Other;

Query Match             3.9%;  Score 66.8;  DB 4;  Length 919;
Best Local Similarity  52.0%;  Pred. No. 2.2e-07;
  Matches  183;  Conservative    0;  Mismatches  157;  Indels   12;  Gaps     1;

Qy           17 AGATGGCGTCCCTGGATGCGGGTGAAGGTACTGGTGTGTTGGGAGACTCAGTGTTGGGAAAT 76
                   ||||||||    |||| ||||||| ||| | |||||| |||||  |||| || ||||||
Db           14 AAATGGCGATGAATAATCGAGTGCGAATTGTTGTGGTTGGCGCCGATTCGGGTGTGGGGAAGA 73

Qy           77 CTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACTGTGG 136
                   |||||     |||||||  || ||  || ||||| |||||||  |||||||| || |||
Db           74 CCTCCCTGACGCACCTGATCACCCACAACGAAGCCCTCATCCGGCCCGGCTGGACACCGTGG 133

Qy          137 GCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAAGAACCCCAGAAGAGAAGACCT 196
                   |||||| ||||||  |||||||  ||||||| ||||  || ||||| | |||||| |||
Db          134 GCTGCAACATCCAGGTGAAGATGCATCCGTTCAGGGAGGGCACCGTCGCGGAGTGTCCCT 193

Qy          197 ACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGAAAAGCA 256
                   |||  || | ||||| ||||| ||||| || | ||  |||   |||   -----------AACCACAAGAACA 241
Db          194 ACTTCGTGGAGCGTGTTCGATGTTGGGGGATCGCTG------------AACCACAAGAACA 241

Qy          257 CAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGACTTAACAAATA 316
                   | |||| ||| || |    |  || |||||| ||| |||| || || ||||||| ||
Db          242 CGCGCAGCGTCTTCTTCATGCGGGCATCGATGGAATCATTCTCGGTGCAGACCTTACCAACG 301

Qy          317 AGAAGTGCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGA 368
                   ||||
Db          302 CCAAGTCGCAGAGGCAGCAGCTAATCGACTGGCTTGTACCAGAGATCGTCAACAAGGA 353
```

```
XX
DB   Human spliced transcript detection oligonucleotide SEQ ID NO:10523.
XX
KW   Human; mouse; rat; splice transcript; detection; RNA transcript;
KW   splice variant; transcriptome; oligonucleotide library; ss.
XX
OS   Homo sapiens.
XX
PN   WO200210449-A2.
XX
PD   07-FEB-2002.
XX
PF   20-JUL-2001; 2001WO-IB001903.
XX
PR   28-JUL-2000; 2000US-0221607P.
PR   02-MAY-2001; 2001US-0287724P.
XX
PA   (COMP-) COMPUGEN INC.
XX
PI   Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX
DR   WPI; 2002-257383/30.
XX
PT   New oligonucleotide libraries comprising oligonucleotides which
PT   selectively hybridise to mRNAs transcribed from a transcription unit of a
PT   genome, useful for detecting tissue-, pathology-, and developmental-
PT   specific genes.
XX
PS   Example 1; SEQ ID NO 10523; 47pp; English.
XX
CC   The present invention describes oligonucleotide libraries for detecting
CC   messenger RNAs that populate a (sub-)transcriptome, where the (sub-
CC   )transcriptome comprises messenger RNAs transcribed from multiple
CC   transcription units that populate a genome. The library comprises several
CC   oligonucleotides, each capable of hybridising selectively to a set of
CC   messenger RNAs transcribed from a given transcription unit of the genome,
CC   which encodes one or more messenger RNA splice variants. The
CC   oligonucleotide libraries are useful for detecting mRNAs from a
CC   biological sample, in expression profiling studies, in qualitatively or
CC   quantitatively characterising the corresponding transcriptome, and in
CC   detecting RNA transcripts and splice variants of human or animal
CC   transcriptomes. The libraries may also be used as specialised mini
CC   libraries to detect transcripts of a sub-transcriptome under a particular
CC   biological or pathological state, and so allowing the detection of tissue
CC   - and pathology-specific genes such as those genes only expressed in
CC   specific tissue under a specific pathological condition; to detect
CC   developmental specific genes; and to detect RNA transcripts and splice
CC   variants of a transcriptome of a patient suffering from a particular
CC   disorder. ABN27253 to ABN59589 represent oligonucleotide sequences from
CC   rats, humans and mice, which are used in the exemplification of the
CC   present invention. N.B. The sequence data for this patent did not form
CC   part of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 60 BP; 7 A; 16 C; 13 G; 24 T; 0 U; 0 Other;

Query Match             3.5%;  Score 60;  DB 6;  Length 60;
Best Local Similarity 100.0%;  Pred. No. 4.6e-06;
  Matches   60;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps     0;

Qy          891 GGGAGGTCCAAACTTTGTCCCTGTTCCTTACCTTTCTGTGTCCCTGTGTATAG 950
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db            1 GGGAGGTCCAAACTTTGTCCCTGTTCCTTACCTTTCTGTGTCCCTGTGTATAG 60
```

```
Search completed: June 30, 2005, 14:23:50
Job time : 1027.81 secs
```

```
RESULT 15
ABN37775
ID   ABN37775 standard; DNA; 60 BP.
XX
AC   ABN37775;
XX
DT   15-JUL-2002  (first entry)
```

Copyright (c) 1993 - 2005 Compugen Ltd.
GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: June 30, 2005, 13:44:48 ; Search time 317.669 Seconds
(without alignments)
8725.600 Million cell updates/sec

Title: US-09-945-173-1
Perfect score: 1694
Sequence: 1 agggaaggcagtggcaagat.........aaacatgaattattgaacta 1694

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :   Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	% Score	Query Match	Match %	DB	Length	ID	Description
1	321	18.9	321	4	321	US-09-513-999C-22021	Sequence 22021, A
2	66.8	3.9	1087	4	1087	US-09-270-767-13908	Sequence 13908, A
3	51.8	3.1	1141	4	1141	US-09-806-708B-22	Sequence 22, Appl
c 4	42.8	2.5	832	4	832	US-09-621-976-2813	Sequence 2813, Ap
c 5	42.8	2.5	1141	4	1141	US-09-806-708B-22	Sequence 22, Appl
c 6	42	2.5	601	4	601	US-09-949-016-51128	Sequence 51128, A
c 7	41.8	2.5	832	4	832	US-09-621-976-2813	Sequence 2813, Ap
8	41.2	2.4	2239	4	2239	US-09-710-279-4411	Sequence 4411, Ap
9	41.2	2.4	2259	4	2259	US-09-710-279-4078	Sequence 4078, Ap
c 10	41.2	2.4	3475	4	3475	US-09-710-279-4032	Sequence 4032, Ap
c 11	40.8	2.4	118923	4	118923	US-09-949-016-13227	Sequence 13227, A
12	40.6	2.4	264665	4	264665	US-09-949-016-13747	Sequence 13747, A
13	39.8	2.3	2273	4	2273	US-08-714-918-40	Sequence 40, Appl
14	39.8	2.3	2273	3	2273	US-09-265-315-40	Sequence 40, Appl
15	39.8	2.3	2273	3	2273	US-09-265-315-40	Sequence 40, Appl
16	39.8	2.3	2273	3	2273	US-09-266-417-40	Sequence 40, Appl
17	39.8	2.3	2273	4	2273	US-09-528-709-40	Sequence 40, Appl
18	39.6	2.3	2273	4	2273	US-09-527-745-40	Sequence 40, Appl
c 19	39.6	2.3	601	4	601	US-09-949-016-130492	Sequence 130492,
20	39.6	2.3	3001	4	3001	US-09-539-333D-170	Sequence 170, App
21	39.6	2.3	197875	4	197875	US-09-949-016-15425	Sequence 15425, A
22	39.2	2.3	40465	4	40465	US-09-949-016-12561	Sequence 12561, A
23	38.4	2.3	300598	4	300598	US-09-949-016-11868	Sequence 11868, A
24	38.4	2.3	302604	4	302604	US-09-949-016-14588	Sequence 14588, A
25	38.4	2.3	302604	4	302604	US-09-949-016-14589	Sequence 14589, A
26	38.4	2.3	308362	4	308362	US-09-949-016-17119	Sequence 17119, A
c 27	38	2.2	601	4	601	US-09-949-016-206666	Sequence 206666,
c 28	38	2.2	64024	4		US-09-949-016-17593	Sequence 17593, A
29	37.6	2.2	601	4		US-09-949-016-56497	Sequence 56497, A
c 30	37.6	2.2	601	4		US-09-949-016-206667	Sequence 206667,
c 31	37.6	2.2	7218	1		US-08-232-463-14	Sequence 14, Appl
c 32	37.6	2.2	136058	4		US-09-949-016-12565	Sequence 12565, A
c 33	37.6	2.2	136480	4		US-09-949-016-17064	Sequence 17064, A
34	37.6	2.2	636591	4		US-09-949-016-11808	Sequence 11808, A
35	37.6	2.2	636591	4		US-09-949-016-13388	Sequence 13388, A
c 36	37.4	2.2	601	4		US-09-949-016-168805	Sequence 168805,
37	37.4	2.2	6722	4		US-09-949-016-13258	Sequence 13258, A
38	37.4	2.2	157032	4		US-09-949-016-16502	Sequence 16502, A
c 39	37.2	2.2	670689	4		US-09-949-016-12505	Sequence 12505, A
c 40	37.2	2.2	670690	4		US-09-949-016-14207	Sequence 14207, A
41	37	2.2	601	4		US-09-949-016-94845	Sequence 94845, A
42	37	2.2	601	4		US-09-949-016-94846	Sequence 94846, A
c 43	37	2.2	9636	4		US-08-323-170B-1	Sequence 1, Appli
c 44	37	2.2	9636	3		US-08-954-441-1	Sequence 1, Appli
c 45	37	2.2	19861	4		US-09-949-016-14498	Sequence 14498, A

ALIGNMENTS

RESULT 1
US-09-513-999C-22021
; Sequence 22021, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 22021
; LENGTH: 321
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-513-999C-22021

Query Match   18.9%; Score 321; DB 4; Length 321;
Best Local Similarity   100.0%; Pred. No. 4e-86;
Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   871 ATGAAAGGAAGTGACAATACGGAGGTCCAAACTTGTCCCTGTTCTGTGTTCCTTAC 930
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATGAAAGGAAGTGACAATACGGAGGTCCAAACTTGTCCCTGTTCTGTGTTCCTTAC 60

Qy   931 CTTTCTGTCCCTGTGTATAGATTATGTGTAAATATGAGATGTTGTCAAA 990
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CTTTCTGTCCCTGTGTATAGATTATGTGTAAATATGAGATGTTGTCAAA 120

Qy   991 ATGATGCAGTAAATGAGCAATGACAGTGTACTGCAGAGAAAATTTACTCTTGCCTAGAAC 1050
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ATGATGCAGTAAATGAGCAATGACAGTGTACTGCAGAGAAAATTTACTCTTGCCTAGAAC 180

Qy   1051 TGGAGGGTTTTTATGGGGTCGTAATTTTCCCACACTCATTGCTGAAAGCTTAATTAAGTA 1110
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TGGAGGGTTTTTATGGGGTCGTAATTTTCCCACACTCATTGCTGAAAGCTTAATTAAGTA 240

Qy   1111 CTTCAAAAACGTATCTCCATTGTTTTACCTTCTTGAGGGGAACGGTCTTGTTAACCAGCC 1170
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CTTCAAAAACGTATCTCCATTGTTTTACCTTCTTGAGGGGAACGGTCTTGTTAACCAGCC 300

Qy   1171 CTGAGTTGTCTACCCCCAAACA 1191
            ||||||||||||||||||||||
Db   301 CTGAGTTGTCTACCCCCAAACA 321

RESULT 2
US-09-270-767-13908
; Sequence 13908, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13908
; LENGTH: 1087
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-13908

Query Match   3.9%;   Score 66.8; DB 4; Length 1087;
Best Local Similarity   52.0%; Pred. No. 2.2e-09;
Matches   183; Conservative   0; Mismatches 157; Indels 12; Gaps 1;

Qy   17 AGATGGCGTCCCTGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAAT 76
             |||||| | ||||||| |||
Db   267 AAATGCGGATGATGAATAATCGAGTGCGAATTGTTGTGTGGTTGCGAATTCGGGTGTGGGGAAGA 326

Qy   77 CTTGTTGGTTAGTCCCATCTCCTATGCCAAAATCAAGTGCTCGGAAAATCCATCATGGACTGTGG 136
             || | || |||| ||||| ||||| ||||||||||
Db   327 CCTCCCTGACGCACCTGATCACCCACCACGAAGCCCTCATCGGGCCCGGCCGGCTGGACGGTGG 386

Qy   137 GCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAAGAACCCCAGAAGAGAGAGACCT 196
             ||||||| ||| || |||||||||| ||||||||
Db   387 GCTGCAACATCCAGTGAAGATGCATCCGTTCAGGGAGGCACCGTCGCGAGTGTCCCT 446

Qy   197 ACTACATAGAGATTATTATGGGATGTTGGAGGCTTCGTGTGGGAGGCTGCCAGCAGCGTGAAAGCA 256
             ||||||| ||||| |||||||
Db   447 ACTTCGTGGGAGCTGTTCGATGTTGGGGGATCGCTG--------AACCACAAGAACA 494

Qy   257 CAAGAGCAGTATTCTACAACTCCGTAAAATGGTATTATTTCGTACACGACTTAACAAATA 316
             |||||| || |||||||||||
Db   495 CGGCGCAGCCGTCTTCATGCGGGCATCGATGAATCATTCGTGTGCACGACCTTACCAACG 554

Qy   317 AGAAGTCCTCCCAAAACTTGCGTCGTTCGGTCATTCGAAGCTCTCAACAGGGA 368
             |||||| ||||
Db   555 CCAAGTCGCAGAGAGCAGCTAATCGACTGGCTGTACGAGATCGTCAACAAGGA 606

RESULT 3
US-09-806-708B-22
; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match   3.1%;   Score 51.8; DB 4; Length 1141;
Best Local Similarity   11.1%; Pred. No. 7.2e-05;
Matches   84; Conservative 287; Mismatches 382; Indels 7; Gaps 2;

Qy   860 TTAATGGAAAAATGAAAGGAAGTGACAATACGGAGGTCCAAACTTTGTCCCTGTTCTCT 919
                   :::::: : ::::: : :|::.::| : ||:|.:|::.:::::
Db   103 WKDMKTAYBMTMTNKWGKTGWRHRYWRWRAMBDTVDHHYVTAMNNAWTTMCMMDKDDKRT 162

Qy   920 GTGTTCCTTACCTTTCTGTCCCTGTGTATAGATTATGTATGTAAAAGCCTTGTGTAAAATATGAG 979
                :: ::| :|||::: :|:||:|:::::
Db   163 RWWWKKNNNATGWDDDTKYHWWNNNGCBTVTWMVRYKTDRDWSBKRMNYGMBWWKNWSYD 222

Qy   980 ATGTTGTCAAAT-GATGCAGTAAATGAGCAATGACAGTGTACTGCGAGAGAAAATTTACT 1038
             :|:: |||||||||::|::::|::|:||||: :::|:.::::
Db   223 VTYYWWVWDDMCKRKVRRWVRTRGRMRNYMVAWBTAHRRRYNNGWTBAMAYRRWTMNNNN 282

Qy   1039 CTTGCCTAGAACTGGAGGGTTTTTATGGTCTGTAATTTTCCCACACTCATTGCTGAAAG 1098
                :.: :|| ||| :::: :::::: ::|:|::   ::: :: ::: ::: :::
Db   283 NNAKAMCKRAKYWGWNRABVNSTCTTWKSKTTKVRTSCWANNCRAGDANKDHKWWKWSAA 342

Qy   1099 CTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTACCTTCTTGAGGGGAACGGTCT 1158
                :. : : |:::|:::|::::   ::   : :   : |||::::::: :::
Db   343 MGVYWNNNNNNNNWTYKKARHBARWDWWHSAWKKWHANAAHYSRKKWTBYKRKTMVNNNN 402

Qy   1159 TGTTAACCAGCCTCAGTTGTCTCCACCCCAAACAATCTCTGTCATTTTCAAAGATGCAAA 1218
             |::::| | || ::::| : : :   :::: :: :::::::::::::::
Db   403 GTTMWKRMWAWYWKMDMDWBGTYNMNNNGGRTYYGWTNKRKWWTYYKWKANNCKKRAWDH 462

Qy   1219 TGGTGTTATTTAATTGTCTCCACCATTGTCACACACAGGAATGCCTAATAATAGCAACCC 1278
                ::: ::: :::: ::: :: ::||: :   -----SHRBAAAVYTWYMWWRRYAHNNN
Db   463 KTCTHNNTTWWKMKTYWNNCYWKSMTNGK-----SHRBAAAVYTWYMWWRRYAHNNN 516

Qy   1279 TTGTCTCCCTCTTCTCTCCTTGCAAATGGCTCAGTGACTCGAAGAGGCGGACTAATAGC 1338
                :: ::::::::|::: |:::: :: : |:::::::: ::::: :::
Db   517 NWDYWWKACTWYKYBVCSKWWNNYAAWYTKSSWNYTSRYYRWKTNNSWRWRSDTRSMGRA 576

Qy   1339 CAGAGTTAAATATAAATACAAATTAATAATATCATAGAGAACAATACCAGAAAAAAAG 1398
               :| : | :|::::: ::::|: |:   :::::::| ::::::: ::::::
Db   577 NNYARABHYGYKWNTRWWBWSHTWBHBRAGAAHYWMBMMYBAKCHCMKAWYKAKKYAGAG 636

Qy   1399 AATTCTGGTAAAATGATGTGAAAAAATTGACAGCTCCCTCACTCTTAAGGTTGCTGTCTATA 1458
                 :: :::::::::::|:::: ::::::: :::: :::::::::::
Db   637 GSNNNNNNNNNNNNNNNATCARDDYYAASRWYAMANAKWYYKBAANNAYYTHANNWWG 696

Qy   1459 TACAGTCTAGGTTTCTGTTTGGAAATAGGTAGGGTAAAATCTAAGACCTGCACAAGGGC 1518
              |||: :|: ::|::: |::::::::::::::|:|:: |||:::
Db   697 CWNNATDTRRTMWKNNNNNNAGTWKNNNNNNAKNASAAKNYAAAAVKCAAKKHWRWANKWA 756

Qy   1519 AGTGAGAGACATTTACAGCCTCCTCCTCTCTATTTTGTTTTTTAAGGAAAAGTCAACTCCTGA 1578
              :: :: :::::::|:::|::::|: ::::: ||||:::
Db   757 MRGWHADAAABTTDKRNNGAYTKYTTTNNNNTYRGVVTNTAARDGWANNNNNNNNNNN 816

Qy   1579 AATGTCCCTTAGCTATAATCAGAAAAACTAAGAAATATTATT 1618
              :: ::::::|:: :|::|:::
Db   817 NGWSDMWVTWWAYANYGTNNNNNNNNNNAYAWWTNKWYYT 856

RESULT 4
US-09-621-976-2813
; Sequence 2813, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2813
; LENGTH: 832

```
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: 235..399
US-09-621-976-2813

Query Match        2.5%;   Score 42.8;   DB 4;   Length 832;
Best Local Similarity   12.0%;   Pred. No. 0.03;
Matches   43;   Conservative 166;   Mismatches 148;   Indels   1;   Gaps   1;

Qy  1028  GAAAATTTACTCTTGCCTAGAACTGGAGGGTTTTTATGGGTCGTAATTTTCCCACACTC  1087
          ::  |::  ::::: :::  :  ::::::  | :  ::::|||:|  : :::
Db    18  KWKNSWSYWMWYWKWYYMKTYWRWRRKKKKCKAWWKYWKITWTWYWRYANWGTYKKKAMCRTK  77

Qy  1088  ATTGCTAAAAGCTTAAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTACCTTCTTGAG  1147
          :::  :::  |||:  :  ::::  :::  :  |::  |:|
Db    78  TKKKKKKGYMWMWYWGWRRSYMAMWTRTWTGYAYYRSMYWWRYRCWKKKAYYRKTTCYS  137

Qy  1148  GGGAACGGTCTGTGTTAACCAGCCTGAGTTGTCTCACCCCAAACAATCTCTGTCATTTTCA  1207
          ::|  |:  :::::::  |:  ::  :::||:::||||:|::::|:::::::
Db   138  SKGWTWWKRWKKAWTTWWWKKTYYWAATRYWWMCWTKRWRASWWYCWWHGKARKWSTWR  197

Qy  1208  AAGATGCAAAATGTGTGTATTTAATTGTCTCCACCATTGTCACACAGGAATGCCTAAT  1267
          ::  ::  :::::  :  ::::  |:  :::  |::  :::  |:|
Db   198  KSRSYASARSA-KRCCYSCSWGAMSWKYWRWRWRGWATGAGMKAWRASCMRRRKYAGK  256

Qy  1268  AATAGACAACCCTGTCTCCCTGTCTCCCTTTGCAAATGGCTCAGTGACTGGAAGAGGC  1327
          ::|  :  ::::::  :  ::::  :  ::::::
Db   257  SKTSYKSMWMMCWTRSWKYCYTKARWTGYYCYRKGGMWCKGRWYASKKYMWKRWWWCWAR  316

Qy  1328  GGACTAATAGCCAGAGTTAAATATAAATACAAATTAATAATAATACATAGAGAACAGCAAT  1385
          |||  :::  :  ::::  :  ::::::  :|:::|:|||  |||
Db   317  MYRYSTGTRASMWWRRWYYTMMWKWYAWARAAWRWAMWAWRRACAAAATATAT  374

RESULT 5
US-09-806-708B-22/c
;  Sequence 22, Application US/09806708B
;  Patent No. 6784342
;  GENERAL INFORMATION:
;  APPLICANT: The University of British Columbia
;  TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
;  FILE REFERENCE: 4810-58741
;  CURRENT APPLICATION NUMBER: US/09/806,708B
;  CURRENT FILING DATE: 2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/147,133
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.0
;  SEQ ID NO 22
;  LENGTH: 1141
;  TYPE: DNA
;  ORGANISM: Artificial sequence
;  FEATURE:
;  NAME/KEY: promoter
;  LOCATION: (1)..(1141)
;  OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match        2.5%;   Score 42.8;   DB 4;   Length 1141;
Best Local Similarity   10.7%;   Pred. No. 0.037;
Matches   90;   Conservative 298;   Mismatches 441;   Indels   13;   Gaps   1;

Qy   833  TAAAATGCTACCCTTCAGCCTTACCCTTTAATCGAAAAATGAAAAGGAAGTGACAAATACGG  892
          :  |::  :|  ::::|::  ::||:  :|:|:|
Db   984  KWARWAGNMRMRMYGAAAGNKWGCMAAAMATMGBWWADTAGKMCNNNNNWTTDVRRMAMKA  925

Qy   893  GAGGTCCAAACTTTGTCCCTGTCTCTGTGTTCCTTACCTTTCGTCCCTGTGTATAGAT  952
          :  :  ::||  :::  ::|:  |
Db   924  KNNNNNNAYWTACYNRAATNNKMATHWMKWTHGAHSKRRTRIHHTRTCRRTKYNNNNNNAR  865

Qy   953  TATGTAAAAGCCTTGTGTAAAATATGAGATATGTGTCAAAATGATGCAGTAAATGAGCAATG  1012
```

```
          |::  ::::||:  :|:  ::
Db   864  TVYWYHHAARRWMNAWWTRTNNNNNNACRNTRTWWABWKHSWCNNNNNNNNNNN  805

Qy  1013  ACAGTGTACTGCAGAGAAAATTACTCTTGCCTAGAACTGGAGGGTTTTTATGGGTCTGT  1072
          :  ::|:  :  :  ::::  ::||
Db   804  NTWCHYTTANABBCYRANNNNAAARMARTCNNNYMHAAVTTHTDWCYKTWMNTWYWDMT  745

Qy  1073  AATTTCCCACACTCATTGCTGAAAGCTTAATTAAGTACTTCAAAAACGTATCTCCATTG  1132
          :  ||:::::  :  :  :::  ||  |
Db   744  TMBTTTTRNMTTSTNMTNNNNNNWACTNNNNNNMWKAYYAHATNNWGCWWNNTDARRTN  685

Qy  1133  TTTTACCTTCTTGAGGGGACGGTCTTGTTGTCTCACCAGCCTGAGTTGTCTACCCCAAACAA  1192
          ||:|  ::::  :  :  :  :::::||
Db   684  NTTVMRRRWMTNTKTRWYSTTRRHHYTGATNNNNNNNNNNNNNSCCTCTRMMTMRWT  625

Qy  1193  TCTCTGTCATTTTCAAAAGATGCAAAATGTGTTATTTAATTGTCTCCACCATTGTCACAC  1252
          :::  ::::  ::  :|:|  :::|  :  :::
Db   624  MKGDGMTVRKKVKWRDTTCTYVDVWADSWVWWYANWMRCRDVTYTRNNTYCKSYAHSYWY  565

Qy  1253  ACAGGAATGCCTAATAATAGCACAACCCATCTCTGTCATTTTCA  1312
          ::  :  :  ::  :|:||:  :  ||:  :  ::::::||:  |
Db   564  WSNNAMWTRRYSARNWSSMARWTTRNNWMSGBVRMRWAGTMWWRHWNNNTDTRYYWWW  505

Qy  1313  GTGACTGGAAGAGGCGGACTAATAGCCAGAGTTAAATATAAATACAAATTAATAATACAT  1372
          :::  ::  :  :::  :
Db   504  KRWA----------RBTTTVVDSMCNAKSMWRGNNWRAMKWWAANNDAGAMDHWTY  458

Qy  1373  AGAGAACAGCAATACCAGAAAAAAGAATTCTGTGAAAAATTGACAGCT  1432
          ::|:  :::  ::  :  ::::  :  :::::
Db   457  WMGNNTMWMRRAWKWMNNAWCRRAYCNNNNNRACVWHKHKMWRWTKYMWKAACNNNNB  398

Qy  1433  CCCTCACTCTTAAGGTTGCTGCTCGTATATACAGTCAGGTTTTCTGTTGGAAATAGGTAGG  1492
          :::  |||:  :::  :  :  ::  ::|  ::|:  ::::
Db   397  KAMYMRVAWMMYSRDTTNTDWMWMTSDWBWHWTYVDYTMRAWNNNNNNWRBCKTTSWM  338

Qy  1493  GTAAAATCTAAGACCTGCACAAGGGCAGTGAGAGACATTTACAGCCTCCTCCTCATTTGT  1552
          ::  :|  |||:  :  ::|::::
Db   337  WWMDHMNTHCTYGNNTWGGSAYBWMAAMSMWAAGASNBVTYNWCWRMTYMGKTMTNNNNNK  278

Qy  1553  TTTTTTAAGGAAAAGTCAACTCCTGAAATGTCCCTTAGCTATAATCAGAAAACTAAGAAT  1612
          :  ::  :  ::::  :
Db   277  AWYRTKTVAWCNNRYYDTAVWTBKRNYKYCYAYBWYBYMYGKGHHWBWWRRABHRSWN  218

Qy  1613  ATTATTCTGTGTCAACAATGTATTTATGGAGAGAAGTAAAAATAAGTTCCACGCAACAC  1672
          ::::  :  :::|:  :::  ::  |::::
Db   217  MWVKCRNKYMVSWHYHAMRYBKWABAVGCNNNWKDRMAHHHWCATNNNMWWWYAYMHH  158

Qy  1673  AA 1674
          ::
Db   157  MH 156
```

```
RESULT 6
US-09-949-016-51128/c
;  Sequence 51128, Application US/09949016
;  Patent No. 6812339
;  GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:   WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE: 2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 51128
;  LENGTH: 601
```

```
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-51128

Query Match        2.5%;  Score 42;  DB 4;  Length 601;
Best Local Similarity  52.3%;  Pred. No. 0.043;
Matches  90;  Conservative  1;  Mismatches  81;  Indels  0;  Gaps  0;

Qy   1520  GTGAGAGACATTTACAGCCCTCCTCCTCTATTTGTTTTTTTAAGGAAAAGTCAACTCCTGAA  1579
                 ||||  ||  ||  ||||||||||||||||||||||||||  |||||||||||||||||
Db    340  GTGATCCACCATGCCCAGCCTGACTTGGTCTTTTTTTTTTTTTAAAAAAAAAAAAGTCTTCAA   281

Qy   1580  ATGTCCCTTAGCTATAATCAGAAAACTAAGAATATTATTCTGTGTCAACAATGTATTTAT  1639
                 ||||||||  ||||||||  ||  || || ||   |||||||  ||||||
Db    280  GTGATTCTGATTCATAATCAGGTTAAGAAGTATCTTAAGGATAATTACAATGTTTATAA   221

Qy   1640  GGAGAGAAGTAAAAAATAAGTTCCACAGCAACACAAAAACATGAATTATTGAA  1691
                 |   ||||||    |||||||
Db    220  TAATAAAAAAATAAGCTACTCTTACAGCTTAGAAATGAAGCAGGAAA  169

RESULT 7
US-09-621-976-2813/c
; Sequence 2813, Application US/09521976
; Patent No. 6639063
; GENERAL INFORMATION:
;   APPLICANT: Dumas Milne Edwards, J.B.
;   APPLICANT: Jobert, S.
;   APPLICANT: Giordano, J.Y.
;   TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;   FILE REFERENCE: GENSET.054PR2
;   CURRENT APPLICATION NUMBER: US/09/621,976
;   CURRENT FILING DATE:  2000-07-21
;   NUMBER OF SEQ ID NOS:  19335
;   SOFTWARE: Patent.pm
; SEQ ID NO 2813
;   LENGTH: 832
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 235..399
US-09-621-976-2813

Query Match        2.5%;  Score 41.8;  DB 4;  Length 832;
Best Local Similarity  11.9%;  Pred. No. 0.06;
Matches  43;  Conservative  167;  Mismatches  149;  Indels  2;  Gaps  1;

Qy    923  TTCCTTACCTTTCTGTCCCCTGTGTATAGATTATGTAAAGCCTTGTGTAAATATGAGATG   982
               :|:||||:||:|: ::|  : :::|  : :::||:   :::::::::
Db    360  WTWWKTWWYWTTYTWTRMWWMKKKARRWYYWWKSTYACASRYRKYTWGWWWYMWKRMST   301

Qy    983  TTGTCAAATGATGCAGTAAATGAGCAATGACAGTGTACTGCGAGAAATTTACTCCTTG  1042
               :::|| :::: :|: :::: : ::|::::|::|: :::|:::::|::|:::|
Db    300  RWYCYMCWKCCMYRGRRCAWYTMARGRMWSYAWGKWKSMRSAMSMCTRMYYKKGSTYWTM   241

Qy   1043  CCTAGAACTGGAGGGTTTTTATGGGTCTGTAATTTTCCCACACTCATTGCTGAAAGCTTA  1102
               :|| ::: ::::|::|::|::::| :|:||::::||:::::::||:||:::||
Db    240  KCTCATWCYWYWKYWKRMWSKTCWSGSRGGYMTSYTSTRSYSMYWASWMYTMCWWWGRWW   181

Qy   1103  ATTAAGTACTTCAAAAACGTATCTCCATTGTTTTTACCTTCTTGACGGGAACGGTCTTGTT  1162
               |:: :::::|: : :::::|  :|::::  |:|::|: :::::::::||
Db    180  STYWYMAWGKKWWRYA--TTWWRRAMMWWWAAWTMWWYMWWWYMMWWAWCMSSRGRAMYRRTMMMWWG  123

Qy   1163  AACCAGCCCTGAGTTGTCTACCCCAAACAATCTCTGTCATTTTCAAAGATGCAAAATGGT  1222
               || :::::::|  ::::|:  ::  :: ::: :: :|:| :::::|
Db    122  YRYWWRKKSYRRTRCAWAYAWKTKRSYYWCWRWKWKRCMWMMMWWAMAYGKTMMRACWKT   63

Qy   1223  GTTATTTAATTGTCTCCACCATTGTCACACAGGAATGCCTAATAAATAGCAACCCTTGT  1282
               :::::| ::: :::|: :::: |::|: ::|::|:::  : ::|:: ::
Db     62  RYWRWWAWAMWRMWWTMWWMYYWYWRAMKRRWMWRKWKRSWSWMMWAWGWMTRWAARMWWRW   3

Qy   1283  C  1283
```

At top right:

```
Db         2 Y 2

RESULT 8
US-09-710-279-4411
; Sequence 4411, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;   APPLICANT: KIMBERLY, WILLIAM JOHN
;   TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;   FILE REFERENCE: PU3480US
;   CURRENT APPLICATION NUMBER: US/09/710,279
;   CURRENT FILING DATE:  2000-11-09
;   PRIOR APPLICATION NUMBER: 60/164,258
;   PRIOR FILING DATE: 1999-11-09
;   NUMBER OF SEQ ID NOS: 4472
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4411
;   LENGTH: 2239
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-09-710-279-4411

Query Match        2.4%;  Score 41.2;  DB 4;  Length 2239;
Best Local Similarity  50.5%;  Pred. No. 0.17;
Matches  100;  Conservative  0;  Mismatches  98;  Indels  0;  Gaps  0;

Qy   1335  TAGCCAGAGTGAAATATAAATACAAATTAATAATACATAGAGAACAGCAATACCAGAAAA  1394
                 ||||| || ||||||||  ||||| |  |||  |||||| || |||||| ||||
Db    725  TTGCAGGAATTAAAAATCAATCTGTATAAAATAATGCATAAAAAAATTAAAATATTAAAATT   784

Qy   1395  AAAGAATTCTGGTAAAATGATGTGAAAAAATTGACAGCTCCCTCACTCTTAAGGTTGCTGC  1454
                 ||| ||  |||||   |||| ||| ||  ||||  |||| |  |||  ||||| |||
Db    785  TAAAAATTGAGTTAACCTTAATTGATACATTGTGATTATTTTTAAATTTATAGTCGCAAC   844

Qy   1455  TATATACAGTCTAGGTTTTCTGTTTGGAAATAGGTAGGTAAAATCTAAGACCTGCACAA  1514
                 | |||| |  ||||| |  ||  || |||||||||  |||||||
Db    845  TGTTTTAGATCAGATTTTTAATTCAATGAGTTTTATGTCCAAAATTTTAGCATTACTCCTA   904

Qy   1515  GGGCAGTGAGAGACATTT  1532
                 ||||| ||||||||||
Db    905  TAAAGTTAAACGTATTT   922

RESULT 9
US-09-710-279-4078/c
; Sequence 4078, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
;   APPLICANT: KIMBERLY, WILLIAM JOHN
;   TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
;   FILE REFERENCE: PU3480US
;   CURRENT APPLICATION NUMBER: US/09/710,279
;   CURRENT FILING DATE:  2000-11-09
;   PRIOR APPLICATION NUMBER: 60/164,258
;   PRIOR FILING DATE: 1999-11-09
;   NUMBER OF SEQ ID NOS: 4472
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4078
;   LENGTH: 2259
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: synthetic
;   OTHER INFORMATION: nucleic acid sequence
US-09-710-279-4078

Query Match        2.4%;  Score 41.2;  DB 4;  Length 2259;
Best Local Similarity  50.5%;  Pred. No. 0.17;
```

Matches   100; Conservative   0; Mismatches   98; Indels   0; Gaps   0;

Qy   1335 TAGCCAGAGTTAAATATAAATACAAATTAATAATACATAGAGAACAGCAATACCAGAAAA 1394
          |||||| ||||| |||||| |||||| ||||||||
Db   967 TTGCAGGAATTAAAAATCAATCGTGTATAAAAAAATGAAAAATATTAAAATT 908

Qy   1395 AAAGAATTCTGGTAAAATGATGTGAAAAAATTGACAGCTCCCTCACTCTTAAGGTTGCTGC 1454
          |||||| ||||||
Db   907 TAAAAATTGAGTTAACCTTAATTGATACATTGTGATTTATTTTTAAATTTATAGTCGCAAC 848

Qy   1455 TATATACAGTCGAGGTTTCTCGTTTGGAAATAGGTAGGGTAAAATCTAAGACCTGCACAA 1514
          |||||||
Db   847 TGTTTTAGATCAGATTTTTAATTCAATGAGTTTTATGTCCAAAATTTTAGCATTACTCTA 788

Qy   1515 GGGCAGTCGAGAGACATTT 1532
          |||||| ||||
Db   787 TAAAGTTTAAACGTATTT 770

RESULT 10
US-09-710-279-4032/c
; Sequence 4032, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4032
; LENGTH: 3475
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: nucleic acid sequence
US-09-710-279-4032

Query Match   2.4%; Score 41.2; DB 4; Length 3475;
Best Local Similarity   50.5%; Pred. No. 0.22;
Matches   100; Conservative   0; Mismatches   98; Indels   0; Gaps   0;

Qy   1335 TAGCCAGAGTTAAATATAAATACAAATTAATAATACATAGAGAACAGCAATACCAGAAAA 1394
          |||| | |||| || |||||| |||||| ||||||||
Db   713 TTGCAGGAATTAAAAATCAATCGTGTATAAAAAAATGAAAAATATTAAAATT 654

Qy   1395 AAAGAATTCTGGTAAAATGATGTGAAAAAATTGACAGCTCCCTCACTCTTAAGGTTGCTGC 1454
          |||| | |||||
Db   653 TAAAAATTGAGTTAACCTTAATTGATACATTGTGATTTATTTTTAAATTTATAGTCGCAAC 594

Qy   1455 TATATACAGTCGAGGTTTCTCGTTTGGAAATAGGTAGGGTAAAATCTAAGACCTGCACAA 1514
          |||||||
Db   593 TGTTTTAGATCAGATTTTTAATTCAATGAGTTTTATGTCCAAAATTTTAGCATTACTCTA 534

Qy   1515 GGGCAGTCGAGAGACATTT 1532
          |||||| ||||
Db   533 TAAAGTTTAAACGTATTT 516

RESULT 11
US-09-949-016-13227/c
; Sequence 13227, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13227
; LENGTH: 118923
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13227

Query Match   2.4%; Score 40.8; DB 4; Length 118923;
Best Local Similarity   52.3%; Pred. No. 2.5;
Matches   90; Conservative   0; Mismatches   82; Indels   0; Gaps   0;

Qy   1520 GTGAGAGACATTTACAGCCTCCTCTCATTTGTTTTTTTTAAGGAAAAGTCAACTCCTGAA 1579
          |||||| || ||| |||||||||
Db   2330 GTGATCCACCATGCCCAGCCTGACTTTGGTCTTTTTTTTTTAAAAAAAAAGCTCTTCAA 2271

Qy   1580 ATGTCCCTAGCTATAATCAGAAAACTAAGAAATATTCTGTGTCAACAATGTATTTAT 1639
          || | ||| ||||| ||| |||||||| ||
Db   2270 GTGATTCTGATTCATAATCAGGTTAAGAAGTAGTATCTTAAAGGATAATTACAATGTTTATA 2211

Qy   1640 GGAGAGAAGTAAAAATAAGTTCCACAGCAACACAAAAAACAATGAATTATTGAA 1691
          ||
Db   2210 TAATAAAAAAATAAGCTACTCTTACAGCTTAGAAAACGAAATGAAGCAGGAAA 2159

RESULT 12
US-09-949-016-13747
; Sequence 13747, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13747
; LENGTH: 264665
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-13747

Query Match   2.4%; Score 40.6; DB 4; Length 264665;
Best Local Similarity   50.3%; Pred. No. 4.6;
Matches   100; Conservative   0; Mismatches   99; Indels   0; Gaps   0;

Qy   912 TGTTCTCTGTGTTCCTTACCTTTCGTCCCGTGTATAGATTATGTAAAAGCCTTGTGTA 971
          |||| ||| || || |||||
Db 151401 TGGTTTAGGTGTTCCCTACTTAAGTGTACTTGATTACAAAATTTGGTTTTATTATTGAT 151460

Qy   972 AATATGAGATGTTGTCAAAATGATGACGCAGTAAATGAGCAATGACAGTGTACTGCAGAGAA 1031
          ||| ||| ||
Db 151461 TAGAGCAGAAGTTGAAAATATGTTGTGATCACGGAAATAACAGTCTTTAGAATCAAAC 151520

Qy   1032 ATTTACTCTTGCCTAGAACTGGAGGGTTTTATGGGTCTGAATTTTCCCACACTCATTG 1091
          ||| |||| |||
Db 151521 ATGGACTTGAATCTAAGGTGTGACATTTCTTGTGGGAGCGTGGATTACTTAATCTCTTTT 151580

Qy   1092 CTGAAAGCTTAATTAAGTA 1110
           ||| || || |||
Db 151581 TGCTATGTCTGCTTGTGTA 151599

RESULT 13
US-08-714-918-40
; Sequence 40, Application US/08714918
; Patent No. 6037123
; GENERAL INFORMATION:
;   APPLICANT: Benton, Bret
;   APPLICANT: Lee, Ving
;   APPLICANT: Malouin, Francois
;   APPLICANT: Martin, Patrick K.
;   APPLICANT: Schmid, Molly B.
;   APPLICANT: Sun, Dongxu
;   TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS ANTIBACTERIAL
;   TITLE OF INVENTION: TARGET GENES
;   NUMBER OF SEQUENCES: 111
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Lyon & Lyon
;   STREET: 633 West Fifth Street
;   STREET: Suite 4700
;   CITY: Los Angeles
;   STATE: California
;   COUNTRY: U.S.A.
;   ZIP: 90071-2066
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;   MEDIUM TYPE: storage
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: IBM P.C. DOS 5.0
;   SOFTWARE: Word Perfect 5.1
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/714,918
;   FILING DATE: September 13, 1996
;   CLASSIFICATION: 424
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 60/009,102
;   FILING DATE: December 22, 1995
;   APPLICATION NUMBER: 60/003,798
;   FILING DATE: September 15, 1995
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Warburg, Richard J.
;   REGISTRATION NUMBER: 32,327
;   REFERENCE/DOCKET NUMBER: 222/005
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (213) 489-1600
;   TELEFAX: (213) 955-0440
;   TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 40:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2273 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
US-08-714-918-40

Query Match   2.3%; Score 39.8; DB 3; Length 2273;
Best Local Similarity 46.5%; Pred. No. 0.44; Indels   0; Gaps   0;
Matches 125; Conservative   1; Mismatches 143; Indels   0; Gaps   0;

Qy   822 CAAAATCATCTTAAAATGCTACCCTTCAGCCTTAACCCTTTAATGGAAAAAATGAAAGGAAG 881
          |||| |||||| ||| || || || ||||||| | |||||||||| ||||||
Db   1560 CAATATCACCTTTATATTTATCCTAWACTTGATTCATTATTTTATTTATTTAGTAAAAG 1619

Qy   882 TGACAATACGGGAGGTCCAAACTTTGTCCCGTTCCGTTCCTACCTTTCTGTCCC 941
          || || |: ||| || || ||| |||||||| ||||||| ||||||||
Db   1620 ACATCATATTCTAAGTKGCAWACGCATTCGCGTTAAAATTCATTGCAGTCTTTATCTCAC 1679

Qy   942 TGTGTATAGATTATGTAAAAGCCTTGTGTAAAATAATGAGATGTTGTCAAAATGATGCAGTA 1001
          |||||| || |||||||| ||| |||| | ||||||| || |||||| ||||||| |
Db   1680 ATTATTCATATTATGTATAATCTTTATTTGAATTAAATTTATATTTGACTTAACTGATTAGTA 1739

Qy   1002 AATGAGCAATGACACAGTGTACTGCAGAGAAAAATTTACTCTTGCCTAGAACTGGAGGGTTT 1061
           || || || || ||| || || || ||| |||| | |||||||||||||||||
Db   1740 TAAAACTAACTTCGTTTACTTCAAAGTTTAAATCTTATCGAGTGATATTTCAGATTCTT 1799

Qy   1062 TATGGGTCTGTAATTTTCCCACACTCATT 1090
           |||||||| || ||||||| ||| ||||
Db   1800 TATCTTTTTATAAAATAGCCCTACAATTT 1828

RESULT 14
US-09-265-315-40
; Sequence 40, Application US/09265315
; Patent No. 6187541
; GENERAL INFORMATION:
;   APPLICANT: Benton, Bret
;   APPLICANT: Lee, Ving J.
;   APPLICANT: Malouin, Francois
;   APPLICANT: Martin, Patrick K.
;   APPLICANT: Schmid, Molly B.
;   APPLICANT: Sun, Dongxu
;   TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
;   TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
;   TITLE OF INVENTION: TARGET GENES
;   NUMBER OF SEQUENCES: 111
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Lyon & Lyon
;   STREET: 633 West Fifth Street
;   STREET: Suite 4700
;   CITY: Los Angeles
;   STATE: California
;   COUNTRY: U.S.A.
;   ZIP: 90071-2066
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
;   MEDIUM TYPE: storage
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: IBM P.C. DOS 5.0
;   SOFTWARE: Word Perfect 5.1
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/265,315
;   FILING DATE: March 9, 1999
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/714,918
;   FILING DATE: September 13, 1996
;   APPLICATION NUMBER: 60/009,102
;   FILING DATE: December 22, 1995
;   APPLICATION NUMBER: 60/003,798
;   FILING DATE: September 15, 1995
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Warburg, Richard J.
;   REGISTRATION NUMBER: 32,327
;   REFERENCE/DOCKET NUMBER: 240/247
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (213) 489-1600
;   TELEFAX: (213) 955-0440
;   TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 40:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2273 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
US-09-265-315-40

Query Match   2.3%; Score 39.8; DB 3; Length 2273;
Best Local Similarity 46.5%; Pred. No. 0.44; Indels   0; Gaps   0;
Matches 125; Conservative   1; Mismatches 143; Indels   0; Gaps   0;

Qy   822 CAAAATCATCTTAAAATGCTACCCTTCAGCCTTAACCCTTTAATGGAAAAAATGAAAGGAAG 881
          |||| |||||| ||| || || || ||||||| | |||||||||| ||||||
Db   1560 CAATATCACCTTTATATTTATCCTAWACTTGATTCATTATTTTATTTATTTAGTAAAAG 1619

```
Query Match        2.3%;   Score 39.8;   DB 3;   Length 2273;
Best Local Similarity  46.5%;   Pred. No. 0.44;
Matches  125;  Conservative  1;  Mismatches 143;  Indels  0;  Gaps  0;

Qy     822 CAAAATCATCTTAAAATGCTACCCTTCAGCCTTAATGGAAAAATGAAAGGAAG 881
               |||||| ||   |||   |||||||| |||
Db    1560 CAATATCACCTTTATATTATCCTAWACTTGACTTAWACTTTTTATTATTAGTAAAG 1619

Qy     882 TGACAATACGGAGGTCCAAACTTTGTCCCTGTTCTCGTGTTCCTTACCTTTCTCGTCCC 941
               |||   |   |:||||||||||||  ||||||||||||||||||||||||||||
Db    1620 ACATCATATTCTAAGTKGCAWACGCATTCGCGTTAAATTCATTGCAGTCTTTATCTCAC 1679

Qy     942 TGTGTATAGATTATGTAAAAGCCTTGTGTAAATATGAGATGTGTCAAAATGATGCAGTA 1001
               |||||||||||||| ||||     |||||||||  |||||||
Db    1680 ATTATTCATATTATGTATAATCTTTATTTGAATTTATATTTGACTTAACTTGATTAGTA 1739

Qy    1002 AATGAGCAATGACAGTGTACTGCGAGAGAAAATTTACTCTTGCCTAGAACTGGAGGGTTTT 1061
               |||||||  |||||   ||| :||
Db    1740 TAAAACTAACTTCGTTTACTTCCCACACTCATT 1090
          TAAAACTAACTTCGTTTACTTCTAAATCTTAATCTTATCGAGTGATATTTCAGATTCTT 1799

Qy    1062 TATGGGTCTGTAATTTTCCCACACTCATT 1090
               |||    | |||||||
Db    1800 TATCTTTTTATAAAATAGCCCTACAATTT 1828
```

Search completed: June 30, 2005, 17:00:57
Job time : 325.669 secs

```
Qy     882 TGACAATACGGAGGTCCAAACTTTGTCCCTGTTCTCGTGTTCCTTACCTTTCTCGTCCC 941
               |||   |   |:||||||||||||  ||||||||||||||||||||||||||||
Db    1620 ACATCATATTCTAAGTKGCAWACGCATTCGCGTTAAATTCATTGCAGTCTTTATCTCAC 1679

Qy     942 TGTGTATAGATTATGTAAAAGCCTTGTGTAAATATGAGATGTGTCAAAATGATGCAGTA 1001
               |||||||||||||| ||||     |||||||||  |||||||
Db    1680 ATTATTCATATTATGTATAATCTTTATTTGAATTTATATTTGACTTAACTTGATTAGTA 1739

Qy    1002 AATGAGCAATGACAGTGTACTGCGAGAGAAAATTTACTCTTGCCTAGAACTGGAGGGTTTT 1061
               |||||||  |||||   ||| :||
Db    1740 TAAAACTAACTTCGTTTACTTCCCACACTCATT 1090

Qy    1062 TATGGGTCTGTAATTTTCCCACACTCATT 1090
               |||    | |||||||
Db    1800 TATCTTTTTATAAAATAGCCCTACAATTT 1828
```

RESULT 15
US-09-265-315-40
; Sequence 40, Application US/09265315
; Patent No. 6187541
; GENERAL INFORMATION:
; APPLICANT: Benton, Bret
; APPLICANT: Lee, Ving J.
; APPLICANT: Malouin, Francois
; APPLICANT: Martin, Patrick K.
; APPLICANT: Schmid, Molly B.
; APPLICANT: Sun, Dongxu
; TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
; TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
; TITLE OF INVENTION: TARGET GENES
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/265,315
; FILING DATE: March 9, 1999
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/714,918
; FILING DATE: September 13, 1996
; APPLICATION NUMBER: 60/009,102
; FILING DATE: December 22, 1995
; APPLICATION NUMBER: 60/003,798
; FILING DATE: September 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 240/247
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2273 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-265-315-40

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2005, 16:53:46 ; Search time 1172.77 Seconds
                                      (without alignments)
                                      9059.822 Million cell updates/sec

Title: US-09-945-173-1
Perfect score: 1694
Sequence: 1 agggaaggcagtggcaagat.........aaacatgattattgaacta 1694

Scoring table: IDENTITY_NUC
              Gapop 10.0 , Gapext 1.0

Searched: 6313374 seqs, 3136092125 residues

Total number of hits satisfying chosen parameters: 12626748

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Published Applications NA:*
    1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
    2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
    3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
    4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
    5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
    6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
    7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
    8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
    9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
   10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
   11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
   12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
   13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
   14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
   15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
   16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
   17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
   18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
   19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
   20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
   21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
   22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
   23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
   24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
   25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
   26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                        SUMMARIES

Result   %   Query
No.   Score   Match Length DB Length ID   Description
----   -----   ----- ------ -- ------ --   -----------
  1   1694   100.0   1694   9   1694   US-09-945-173-1   Sequence 1, Appli
  2   1664.6   98.3   1694   19   2398   US-10-479-284-34   Sequence 34, Appl
  3   1358.8   80.2   1417   13   1417   US-10-044-090-467   Sequence 467, App
  4   735   43.4   735   17   735   US-10-210-130-125   Sequence 125, App
  5   731.6   43.2   739   17   739   US-10-210-130-127   Sequence 127, App
  6   711   42.0   711   9   711   US-09-945-173-3   Sequence 3, Appli
  7   628.4   37.1   959   9   959   US-09-764-868-321   Sequence 321, App

  8   444   26.2   444   8   504   10 US-09-918-995-23516   Sequence 23516, A
  9   252.8   14.9   387   9   60   US-09-983-965-2369   Sequence 2369, Ap
 10   60   3.5   60   10   60   US-09-908-975-10523   Sequence 10523, A
 11   54.6   3.2   1626   10 1626   US-10-425-115-2369   Sequence 2369, Ap
 12   51.4   3.0   664   20   664   US-10-767-701-25416   Sequence 25416, A
 13   50.2   3.0   1643   19 1643   US-10-425-114-26886   Sequence 26886, A
 14   50.2   3.0   1643   20 1643   US-10-425-115-75255   Sequence 75255, A
 15c   46.2   2.7   1584   19 1584   US-10-437-963-25042   Sequence 25042, A
 16   45.2   2.7   761   19   761   US-10-437-963-650   Sequence 650, App
 17   44   2.6   6713   15 6713   US-10-311-455-1054   Sequence 1054, Ap
 18   44   2.6   6713   15 6713   US-10-240-485-80   Sequence 80, Appl
 19   43.4   2.6   5641   15 5641   US-10-311-455-1369   Sequence 1369, Ap
 20c   42.8   2.5 3673778   16 3673778   US-10-312-841-1   Sequence 1, Appli
 21c   42.4   2.5   444   10   444   US-09-814-353-17561   Sequence 17561, A
 22   42   2.5   42   18 56153   US-10-221-714A-519   Sequence 519, App
 23c   41.2   2.4   13131   18 13131   US-10-240-589C-57   Sequence 57, Appl
 24c   41.2   2.4   56153   18 56153   US-10-221-714A-520   Sequence 520, App
 25c   41   2.4   533   20   533   US-10-425-115-85901   Sequence 85901, A
 26   40.8   2.4   457   10   457   US-09-918-995-36980   Sequence 36980, A
 27   40.8   2.4   495   19   495   US-10-021-322-4579   Sequence 4579, Ap
 28c   40.8   2.4 176001   17 176001   US-10-210-556-27   Sequence 27, Appl
 29c   40.8   2.4 186739   17 186739   US-10-210-556-19   Sequence 19, Appl
 30   40.4   2.4   879   20   879   US-10-425-115-173241   Sequence 173241,
 31   40.4   2.4 120239   19 120239   US-10-322-281-271   Sequence 271, App
 32c   40.4   2.4 1790242 20 1790242   US-10-719-993-6940   Sequence 6940, Ap
 33   40.2   2.4   551   15   551   US-10-425-115-40080   Sequence 40080, A
 34   40.2   2.4   5461   15 5461   US-10-311-455-972   Sequence 972, App
 35   40.2   2.4   5461   19 5461   US-10-433-793-8   Sequence 8, Appli
 36   40.2   2.4 3673778   16 3673778   US-10-312-841-2   Sequence 2, Appli
 37   40   2.4   436   20   436   US-10-425-115-12033   Sequence 12033, A
 38   39.8   2.3   577   19   577   US-10-021-322-162   Sequence 162, App
 39   39.8   2.3   583   19   583   US-10-021-322-6081   Sequence 6081, Ap
 40   39.6   2.3   7498   15 7498   US-10-311-455-230   Sequence 230, App
 41   39.6   2.3   364   10   364   US-09-918-995-18370   Sequence 18370, A
 42   39.6   2.3   608   20   608   US-10-425-115-80992   Sequence 80992, A
 43   39.6   2.3   3001   17 3001   US-10-147-603-170   Sequence 170, App
 44c   39.6   2.3   6531   15 6531   US-10-311-455-614   Sequence 614, App
 45   39.6   2.3 3673778   16 3673778   US-10-312-841-2   Sequence 2, Appli

                        ALIGNMENTS

RESULT 1
US-09-945-173-1
; Sequence 1, Application US/09945173
; Patent No. US20020127568A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 47324, A NOVEL HUMAN G-PROTEIN AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 38155-20035.00
; CURRENT APPLICATION NUMBER: US/09/945,173
; CURRENT FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/229,293
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1694
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19)....(729)
US-09-945-173-1

  Query Match   100.0%; Score 1694; DB 9; Length 1694;
  Best Local Similarity   100.0%; Pred. No. 0;
  Matches 1694; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 AGGGAAGGCAGTGGCAAGATGGCCTCCCTGGATCGGTGAAGGTACTGGTGTTGGGAGAC 60

```
Db              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy     1  AGGGAAGGCAGTGGCAAGATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGAC    60

Qy    61  TCAGGTGTTGGGAAATCTTCGTTAGTCCATCTTCCTAGTCCATCTTCCTAGTCCATCT...  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  TCAGGTGTTGGGAAATCTTCGTTAGTCCATCTTCCTAGTCCATCTTCCTAGTCCATCT...  120

Qy   121  CCATCATGGACGTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAGGAACC   180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  CCATCATGGACGTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAGGAACC   180

Qy   181  CCAGAAGAAGAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCC   240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  CCAGAAGAAGAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCC   240

Qy   241  AGCAGCGTGAAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTA   300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  AGCAGCGTGAAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTA   300

Qy   301  CACGACTTAACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTC   360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  CACGACTTAACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTC   360

Qy   361  AACAAGGGATTGGTGCCAACTGAGTCTTGGTGACAAATGGGAATGATGATCAAGAACAG   420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  AACAAGGGATTGGTGCCAACTGAGTCTTGGTGACAAATGGGAATGATGATCAAGAACAG   420

Qy   421  TTTGCTGATAACCAAATACCACCTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAA   480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  TTTGCTGATAACCAAATACCACCTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAA   480

Qy   481  ACAAAGCGCCATGAAGTGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAA   540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  ACAAAGCGCCATGAAGTGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAA   540

Qy   541  GAAATTAATTTGGACTGCACAAATCCACGGTCAGTCGCAGTTCTTCCCAATGCTGTC       600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  GAAATTAATTTGGACTGCACAAATCCACGGTCAGTCGCAGTTCTTCCCAATGCTGTC       600

Qy   601  AAGCTTCAGTAGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTAAGAGAAGTAAT     660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  AAGCTTCAGTAGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTAAGAGAAGTAAT     660

Qy   661  CAGATTCCAGGCTTTCCTGAATCGGGAAAAGATTTGGGGCAGGAACATTAAAGAGCCTTCAT  720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  CAGATTCCAGGCTTTCCTGAATCGGGAAAAGATTTGGGGCAGGAACATTAAAGAGCCTTCAT  720

Qy   721  TATGACTGAATTACACTCCTTGGAAGAGTGAGCAAGCAGTGAGCAAGCAGTGGCAGAGC   780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721  TATGACTGAATTACACTCCTTGGAAGAGTGAGCAAGCAGTGAGCAAGCAGTGGCAGAGC   780

Qy   781  TCATCTTGCTGTGTCAATTATTACCATCACAGCCTTTAACAAAATCATCTTAAAATGC   840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  TCATCTTGCTGTGTCAATTATTACCATCACAGCCTTTAACAAAATCATCTTAAAATGC   840

Qy   841  TACCCTTCAGCTTACCCTTAATGAAAATGAAAATGAAAGGAGTGCACAATACGGGAGGTCCA   900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841  TACCCTTCAGCTTACCCTTAATGAAAATGAAAATGAAAGGAGTGCACAATACGGGAGGTCCA   900

Qy   901  AACTTTGTCCCGTGTTCTCGTGTTCCTTACCTTTCTGTATAGATTATGTATGTAAA      960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901  AACTTTGTCCCGTGTTCTCGTGTTCCTTACCTTTCTGTATAGATTATGTATGTAAA      960

Qy   961  AGCCTTGTGTAAATATGAGATGTTGTCAAAATGATGCAATGAGCAATGACAGTGTA      1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  AGCCTTGTGTAAATATGAGATGTTGTCAAAATGATGCAATGAGCAATGACAGTGTA      1020

Qy  1021  CTGCAGAGAAAATTTACTCTTGCCTAGAACTGGAGGGTTTTTATGGTCTGTGTAATTTTCC  1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021  CTGCAGAGAAAATTTACTCTTGCCTAGAACTGGAGGGTTTTTATGGTCTGTGTAATTTTCC  1080

Qy  1081  CACACTCATTGCTGAAAGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTACCT  1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081  CACACTCATTGCTGAAAGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTACCT  1140

Qy  1141  TCTTGAGGGGAACGGTCTTGTTAACCAGCCTGAGTTGTCTACCCCAAACAATCTCTGTC   1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141  TCTTGAGGGGAACGGTCTTGTTAACCAGCCTGAGTTGTCTACCCCAAACAATCTCTGTC   1200

Qy  1201  ATTTCAAAGATGCAAAATGGTGTTATTTAATTGTCTCCACCATTGTCACACACAGGAAT  1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201  ATTTCAAAGATGCAAAATGGTGTTATTTAATTGTCTCCACCATTGTCACACACAGGAAT  1260

Qy  1261  GCCTAATAATAGCAACCCTTGTCTCCTCCTCTTTGCAAATGGCTCAGTGACTGG       1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261  GCCTAATAATAGCAACCCTTGTCTCCTCCTCTTTGCAAATGGCTCAGTGACTGG       1320

Qy  1321  AAGAGGCGGACTAATAGCCAGAGTTAAAATAATAACAAAATAATAATAATACATAGAGAACA  1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321  AAGAGGCGGACTAATAGCCAGAGTTAAAATAATAACAAAATAATAATAATACATAGAGAACA  1380

Qy  1381  GCAATACCAGAAAAAAGAATTCTGGTAAAATGATGTGAAAAATTGACAGCTCCCTCACT   1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381  GCAATACCAGAAAAAAGAATTCTGGTAAAATGATGTGAAAAATTGACAGCTCCCTCACT   1440

Qy  1441  CTTAAGGTTGCTGCTATATACAGTCTAGGTTTTCTGTTTGGAAATAGGTAGGGTAAAATC  1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441  CTTAAGGTTGCTGCTATATACAGTCTAGGTTTTCTGTTTGGAAATAGGTAGGGTAAAATC  1500

Qy  1501  TAAGACCTGCACAAGGGCAGTGAGAGACATTACAGCCTCCTCTCATTTGTTTTTAA     1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501  TAAGACCTGCACAAGGGCAGTGAGAGACATTACAGCCTCCTCTCATTTGTTTTTAA     1560

Qy  1561  GGAAAAGTCAACTCCTGAAATGTCCCTTAGCTATAATCAGAAAACTAAGAATATTATTCT  1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561  GGAAAAGTCAACTCCTGAAATGTCCCTTAGCTATAATCAGAAAACTAAGAATATTATTCT  1620

Qy  1621  GTGTCAACAATGTATTTATGGAGAGAAGAAGTAAAAATAAGTTCCACAGCAACACAAAACAT  1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1621  GTGTCAACAATGTATTTATGGAGAGAAGAAGTAAAAATAAGTTCCACAGCAACACAAAACAT  1680

Qy  1681  GAATTATTGAACTA  1694
          ||||||||||||||
Db  1681  GAATTATTGAACTA  1694
```

RESULT 2
US-10-479-284-34
; Sequence 34, Application US/10479284
; Publication No. US20040158039A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION ; YUE, Henry;
; APPLICANT: LEE, Ernestine A. ; BECHA, Shanya D.;
; APPLICANT: BAUGHN, Mariah R. ; YAO, Monique G.;
; APPLICANT: TANG, Y. Tom ; AU-YOUNG, Janice K.;
; APPLICANT: LAL, Preeti G. ; WARREN, Bridget A.;
; APPLICANT: DUGGAN, Brendan M. ; TRAN, Uyen K.;
; APPLICANT: XU, Yuming ; THANGAVELU, Kavitha;
; APPLICANT: RICHARDSON, Thomas W. ; BANDMAN, Olga;
; APPLICANT: JONES, Karen Anne ; YANG, Junming;
; APPLICANT: EMERLING, Brooke M. ; SWARNAKAR, Anita;
; APPLICANT: LUO, Wen ; CHAWLA, Narinder K.;
; APPLICANT: AZIMZAI, Yalda ; KHAN, Farrah A.;
; APPLICANT: LU, Dyung Aina M. ; GRIFFIN, Jennifer A.;
; APPLICANT: LEE, Soo Yeun ; BURFORD, Neil;
; APPLICANT: ELLIOTT, Vicki S. ; HONCHELL, Cynthia D.;
; APPLICANT: HE, Ann ; MASON, Patricia M.;
; APPLICANT: LI, Joana X. ; HAFALIA, April J.A.;
; APPLICANT: GURURAJAN, Rajagopal
; TITLE OF INVENTION: SECRETED PROTEINS
; FILE REFERENCE: PF-0998 USN
; CURRENT APPLICATION NUMBER: US/10/479,284
; CURRENT FILING DATE: 2003-11-24
; PRIOR APPLICATION NUMBER: PCT/US02/16234
; PRIOR FILING DATE: 2002-05-21
; PRIOR APPLICATION NUMBER: US 60/293,728

```
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/297,019
; PRIOR FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/299,297
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/300,537
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: US 60/301,936
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/362,439
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: US 60/363,649
; PRIOR FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 60/366,041
; PRIOR FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PERL Program
; SEQ ID NO 34
; LENGTH: 2398
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 747515CB1
US-10-479-284-34

Query Match      98.3%;   Score 1664.6;   DB 19;   Length 2398;
Best Local Similarity  99.8%;   Pred. No. 0;
Matches 1667;   Conservative  0;   Mismatches  0;   Indels  4;   Gaps  0;

Qy    17  AGATGGCGTCCCTGGATCGGGTGAAGGTGACTGGTGTTGGGAGACTCAGGTGTTGGGAAAT   76
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1  AGATGGCGTCCCTGGATCGGGTGAAGGTGACTGGTGTTGGGAGACTCAGGTGTTGGGAAAT   60

Qy    77  CTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAAATCATCATGGACTGTGG  136
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61  CTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAAATCATCATGGACTGTGG  120

Qy   137  GCTGCTCAGTGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAGAAGACCT  196
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121  GCTGCTCAGTGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAGAAGACCT  180

Qy   197  ACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTTGCCAGCAGCGTGAAAAGCA  256
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181  ACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTTGCCAGCAGCGTGAAAAGCA  240

Qy   257  CAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACGACTTAACAAATA  316
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241  CAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACGACTTAACAAATA  300

Qy   317  AGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGGACAAATGGGATTATGATCAAGAACCAGTTGCTGATAACCAAA  376
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301  AGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGGACAAATGGGATTATGATCAAGAACCAGTTGCTGATAACCAAA  360

Qy   377  CAACTCGGAGTCTTGGTGACAAATGGGACATTACAGCCTCCTCTATTTGTTTTTTAAGGAAAAGTCAACTCCT  436
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361  CAACTCGGAGTCTTGGTGACAAATGGGACATTACAGCCTCCTCTATTTGTTTTTTAAGGAAAAGTCAACTCCT  420

Qy   437  TACCACTGTTGGTAATAGGGACTAAACTGGACCAGAGAATTCATGAAACAAAGCGCCATGAAG  496
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421  TACCACTGTTGGTAATAGGGACTAAACTGGACCAGAGAATTCATGAAACAAAGCGCCATGAAG  480

Qy   497  TTTTAACTAGGACTGCTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACT  556
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481  TTTTAACTAGGACTGCTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACT  540

Qy   557  GCACAAATCCACGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCAGTAGGTTTT  616
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541  GCACAAATCCACGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCAGTAGGTTTT  600

Qy   617  TTGATAAGGTCATAGAGAAGATACTTTTTAAGAGAAGGTAATCAGATTCCAGGCTTTC  676
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601  TTGATAAGGTCATAGAGAAGAAGATACTTTTTAAGAGAAGGTAATCAGATTCAGATTCCAGGCTTTC  660

Qy   677  CTGATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGCCTTCATTATGACTGATTACAC  736
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661  CTGATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGCCTTCATTATGACTGATTACAC  720

Qy   737  TCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTCACAGCTCATCTGCTGTGTTC  796
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721  TCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTCACAGCTCATCTGCTGTGTTC  780

Qy   797  AATTATTACCATCACAGCCTTTAACAAAATCATCTTAAAATGTACCCTTCAGCCTTAC  856
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781  AATTATTACCATCACAGCCTTTAACAAAATCATCTTAAAATGTACCCTTCAGCCTTAC  840

Qy   857  CCCTTAATGGAAAAATGAAAGGAAGTGACAATACGGGAGGTCCAAACTTTGTCCCTGTTC  916
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841  CCCTTAATGGAAAAATGAAAGGAAGTGACAATACGGGAGGTCCAAACTTTGTCCCTGTTC  900

Qy   917  TCTGTGTTCCTTACCTTTCGTCCTGTGTATAGATTATGTAAAAGCCTTGTGTAAAATAT  976
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901  TCTGTGTTCCTTACCTTTCGTCCTGTGTATAGATTATGTAAAAGCCTTGTGTAAAATAT  960

Qy   977  GAGATGTTGTCAAAATGATGCAGTAAATGAGCAATGACAGTGTACTGCAGAGAAAATTTA 1036
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961  GAGATGTTGTCAAAATGATGCAGTAAATGAGCAATGACAGTGTACTGCAGAGAAAATTTA 1020

Qy  1037  CTCTTGCCTAGAACTGGAGGGTTTATGGGTCTGTAATTTCCCACACTCATTGCTGAA 1096
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021  CTCTTGCCTAGAACTGGAGGGTTTTATGGGTCTGTAATTTTCCCACACTCATTGCTGAA 1080

Qy  1097  AGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTACCTTCTTGAGGGGAACGGT 1156
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081  AGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTTACCTTCTTGAGGGGAACGGT 1140

Qy  1157  CTTGTTAACCAGCCCTGAGTTGTCTACCCCAAACAATCTCTGTCATTTTCAAAGATGCAA 1216
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141  CTTGTTAACCAGCCCTGAGTTGTCTACCCCAAACAATCTCTGTCATTTTCAAAGATGCAA 1200

Qy  1217  AATGGTGTTATTTAATTGTCTCCACCATTGTCACACAGGAATGCCTAATAATAGCAAC 1276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201  AATGGTGTTATTTAATTGTCTCCACCATTGTCACACAGGAATGCCTAATAATAGCAAC 1260

Qy  1277  CCCTTGTCTCCCTCCTCCTCCTTTGCAAATGGCTCAGATGGCTCAGTGACTGGAAGAGGCGGACTAATA 1336
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261  CCTTGTCTCCCTCCTCCTCCTTTGCAAATGGTCAGTGACTGGAAGAGGCGGACTAATA 1320

Qy  1337  GCCAGAGTAAATATAAATAATAACAGAATTAATAATACATAGAACAGCAATACCAGAAAAAA 1396
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321  GCCAGAGTAAATATAAATAATAACAGAATTAATAATACATAGAACAGCAATACCAGAAAAAA 1380

Qy  1397  AGAATTCGGTAAAATGATGTGAAAAATGTGAAATGGTTTCTGTTTGGAAATAGGTAGGTTGCTGCTA 1456
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381  AGAATTCGGTAAAATGATGTGAAAAATGTGAAATGGTTTCTGTTTGGAAATAGGTAGGTTGCTGCTA 1440

Qy  1457  TATACAGTCTAGGTTTCTGTTTCGTTTGGTCATTGGGAAATCTTAAGACCTGACAAGG 1516
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441  TATACAGTCTAGGTTTCTGTTTCGTTTGGTCATTGGGAAATCTTAAGACCTGCACAAGG 1500

Qy  1517  GCAGTGAGACATTTACAGCCTCCTCTATTTGTTTTTTAAGGAAAACAAAGTCAACTCCT 1576
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501  GCAGTGAGACATTTACAGCCTCTCTATTTGTTTTTTTAAGGAAAACAAAGTCAACTCCT 1560

Qy  1577  GAAATGTCCCTTAGCTATAAATCAGAAAACTAAGAATATTATTCTGTCAACAATGTATT 1636
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561  GAAATGTCCCTTAGCTATAATCAGAAAACTAAGAATATTATTCTGTCAACAATGTATT 1620

Qy  1637  TATGGAGAGAAGTCATAGAGAAGAGTAATCAACACAAAAACATGAATTAT 1687
          ||||||||||||||||||||||||||||||||||||||||||||||
Db  1621  TATGGAGAGAAGTCATAGAGAAGAGTAATCAACACAACACATTTACATGAATTAT 1671
```

RESULT 3
US-10-044-090-467
; Sequence 467, Application US/10044090

```
; Publication No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 467
; LENGTH: 1417
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020137081A1 1307098.1
US-10-044-090-467

Query Match             80.2%;  Score 1358.8;  DB 13;  Length 1417;
Best Local Similarity   99.9%;
Matches 1360;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;
```

```
Qy     17  AGATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGTGTGTTGGGAAAT   76
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  AGATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGTGTGTTGGGAAAT   60

Qy     77  CTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGCGACTGTGG  136
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  CTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGCGACTGTGG  120

Qy    137  GCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCAGAAGAGAAGACCT   196
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  GCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCAGAAGAGAAGACCT   180

Qy    197  ACTACAATAGAATTATGGGATGTTGGAGGCTCGTGGGGAGTTATTTCGTACACGACTAACGA  256
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  ACTACAATAGAATTATGGGATGTTGGAGGCTCGTGGGGAGTTATTTCGTACACGACTAACGA  240

Qy    257  CAAGAGCAGTATTCCAAAAACTTGCGTCGTTGGTCGTTCATTGGAAGCTCTCAACAGGGA  316
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  CAAGAGCAGTATTCCAAAAACTTGCGTCGTTGGTCGTTCATTGGAAGCTCTCAACAGGGA  300

Qy    317  AGAAGTCCTCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGC  376
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  AGAAGTCCTCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGC  360

Qy    377  CAACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGAACAGTTGCTGATAACCAAA  436
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  CAACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGAACAGTTGCTGATAACCAAA  420

Qy    437  TACCACTGTTGGTAATAGGGATAGGAGAGATACTTTTTAAGAGAAGTAATCAGATTCCAGGCTTTC  496
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  TACCACTGTTGGTAATAGGGATAGGAGAGATACTTTTTAAGAGAAGTAATCAGATTCCAGGCTTTC  480

Qy    497  TTTTAACTAGGACTGCTTCCTCGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACT  556
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  TTTTAACTAGGACTGCTTCCTCGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACT  540

Qy    557  GCACAAATCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTTGTCAAGCTAGGTTTT  616
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  GCACAAATCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTTGTCAAGCTAGGTTTT  600

Qy    617  TTGATAAGGTCATAGAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGTAATCAGATTCCAGGCGCTTTC  676
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  TTGATAAGGTCATAGAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGTAATCAGATTCCAGGCGCTTTC  660

Qy    677  CTGATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGCCTTCATTATGACTGAATTACAC  736
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  CTGATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGCCTTCATTATGACTGAATTACAC  720

Qy    737  TCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTTCACAGCTCATCTTGCTGTGTTC  796
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  TCATCCTTTGGAAGAGTGAGCAAGCAGTGGCAGTTTTTCACAGCTCATCTTGCTGTGTTC  780

Qy    797  AATTATTACCATCACAGCCTTTAACAAAATCATCTTAAAATGCTACCCTTCAGCCTTAC  856
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  AATTATTACCATCACAGCCTTTAACAAAATCATCTTAAAATGCTACCCTTCAGCCTTAC  840

Qy    857  CCTTTAATGAAAAATGAAAGGAAGTGACAATACGGAGGTCCAAACTTGTCCCTGTTC  916
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  CCTTTAATGAAAAATGAAAGGAAGTGACAATACGGAGGTCCAAACTTGTCCCTGTTC  900

Qy    917  TCTGTGTTCCTTACCTTTCTGTCCCCGTGTATAGATTATGTGTAAAAGCCTTGTGTAAATAT  976
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  TCTGTGTTCCTTACCTTTCTGTCCCCGTGTATAGATTATGTGTAAAAGCCTTGTGTAAATAT  960

Qy    977  GAGATGTTGTCAAAATGATGCAGTAAATGAGCAATGACAGTGTACTGCAGAGAAAATTTA  1036
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961  GAGATGTTGTCAAAATGATGCAGTAAATGAGCAATGACAGTGTACTGCAGAGAAAATTTA  1020

Qy   1037  CTCTTGCCTAGAACTGGAGGGTTTTATGGGTCTGTAATTTCCCACACTCATTGCTGAA  1096
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021  CTCTTGCCTAGAACTGGAGGGTTTTATGGGTCTGTAATTTCCCACACTCATTGCTGAA  1080

Qy   1097  AGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTACCTTCTTCTTGAGGGGAACGGT  1156
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081  AGCTTAATTAAGTACTTCAAAAACGTATCTCCATTGTTTACCTTCTTCTTGAGGGGAACGGT  1140

Qy   1157  CTTGTTAACCAGCCCTGAGTTGTCTCACCCCAAACAATCTCTGTCAATTTCAAAGATGCAA  1216
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141  CTTGTTAACCAGCCCTGAGTTGTCTCACCCCATACAATCTCTGTCAATTTCAAAGATGCAA  1200

Qy   1217  AATGGTGTTATTAATTGCTCCACCATTGTCACACACAGGAATGCCTAATAATAGCAAC  1276
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201  AATGGTGTTATTAATTGCTCCACCATTGTCACACACAGGAATGCCTAATAATAGCAAC  1260

Qy   1277  CCTTGTCTCCCTCTTCTCTCCTTTGCAAATGCCTCAGTGACTGGAACGAGAGGCGACTAATA  1336
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261  CCTTGTCTCCCTCTTCTCTCCTTTGCAAATGCCTCAGTGACTGGAACGAGAGGCGACTAATA  1320

Qy   1337  GCCAGAGTTAAATATAAATACAAAATAATAATACAATAATAATACACATAGAGAA  1378
            ||||
Db   1321  GCCAGAGTTAAATATAAATACAAAATAATAATACAATAATAATACACATAGAGAA  1362
```

```
RESULT 4
US-10-210-130-125
; Sequence 125, Application US/10210130
; Publication No. US2004014053A1
GENERAL INFORMATION:
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Patturajan, Meera
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Miller, Charles E.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Li, Li
; APPLICANT: Berghs, Constance
; APPLICANT: Zhong, Mei
; APPLICANT: Casman, Stacie J.
; APPLICANT: Voss, Edward Z.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Smithson, Glennda
; APPLICANT: Ji, Weizhen
; APPLICANT: Gorman, Linda
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Leite, Mario W.
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Anderson, David W.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Khramtsov, Nikolai V.
; APPLICANT: Ort, Tatiana
```

```
;   APPLICANT: Ellerman, Karen
;   APPLICANT: Rastelli, Luca
;   APPLICANT: Agee, Michele L.
;   APPLICANT: Chaudhuri, Amitabha
;   APPLICANT: Chant, John S.
;   APPLICANT: DiPippo, Vincent A.
;   APPLICANT: Edinger, Shlomit R.
;   APPLICANT: Eisen, Andrew J.
;   APPLICANT: Gangolli, Esha A.
;   APPLICANT: Giot, Loic
;   APPLICANT: Ooi, Chean Eng
;   APPLICANT: Rothenberg, Mark E.
;   APPLICANT: Spaderna, Steven K.
;   APPLICANT: Hjalt, Tord
;   APPLICANT: Liu, Xiaohong
;   APPLICANT: Taupier, Raymond J., Jr.
;   APPLICANT: Catterton, Elina
;   APPLICANT: Shenoy, Suresh G.
;   TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;   FILE REFERENCE: 21402-416C (Cura-716 SMT)
;   CURRENT APPLICATION NUMBER: US/10/210,130
;   CURRENT FILING DATE: 2002-08-01
;   PRIOR APPLICATION NUMBER: 60/309,501
;   PRIOR FILING DATE: 2001-08-02
;   PRIOR APPLICATION NUMBER: 60/316,508
;   PRIOR FILING DATE: 2001-08-31
;   PRIOR APPLICATION NUMBER: 60/354,655
;   PRIOR FILING DATE: 2002-02-05
;   PRIOR APPLICATION NUMBER: 60/310,291
;   PRIOR FILING DATE: 2001-08-03
;   PRIOR APPLICATION NUMBER: 60/383,887
;   PRIOR FILING DATE: 2002-05-29
;   PRIOR APPLICATION NUMBER: 60/310,951
;   PRIOR FILING DATE: 2001-08-08
;   PRIOR APPLICATION NUMBER: 60/323,936
;   PRIOR FILING DATE: 2001-09-21
;   PRIOR APPLICATION NUMBER: 60/381,039
;   PRIOR FILING DATE: 2002-05-16
;   PRIOR APPLICATION NUMBER: 60/311,292
;   PRIOR FILING DATE: 2001-08-09
;   PRIOR APPLICATION NUMBER: 60/311,979
;   PRIOR FILING DATE: 2001-08-13
;   Remaining Prior Application data removed - See File Wrapper or PALM.
;   NUMBER OF SEQ ID NOS: 369
;   SOFTWARE: CuraSeqList version 0.1
;
; SEQ ID NO 125
;   LENGTH: 735
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (10)..(717)
US-10-210-130-125

Query Match       43.4%;  Score 735;  DB 17;  Length 735;
Best Local Similarity  100.0%;  Pred. No. 4.7e-192;
Matches 735;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy      10 AGTGGCAAGATGGCGTCCCTGATCGGGTGAAGGTACTGGGTGTTGGGAGACTCAGGTGTT 69
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AGTGGCAAGATGGCGTCCCTGATCGGGTGAAGGTACTGGGTGTTGGGAGACTCAGGTGTT 60

Qy      70 GGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGG 129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGG 120

Qy     130 ACTGTGGGCGTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAACCCCAGAAGAG 189
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 ACTGTGGGCGTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAACCCCAGAAGAG 180

Qy     190 AAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTG 249
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTG 240
```

```
Qy     250 AAAAGCCAAGAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTA 309
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 AAAAGCCAAGAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTA 300

Qy     310 ACAAATAAGAAGTCCTCCCAAAACTTGCTCGTTGGTCATTGGAAGCTCAACAGGGAT 369
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ACAAATAAGAAGTCCTCCCAAAACTTGCTCGTTGGTCATTGGAAGCTCAACAGGGAT 360

Qy     370 TTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGAT 429
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGAT 420

Qy     430 AACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGC 489
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGC 480

Qy     490 CATGAAGTTTAACTAGGACTGCTTCCTGGCTGAGGATTTCAATCCAGAAGAAAATTAAT 549
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CATGAAGTTTAACTAGGACTGCTTCCTGGCTGAGGATTTCAATCCAGAAGAAAATTAAT 540

Qy     550 TTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCTTCCCAATGCGTGTCAAGCTCAGT 609
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 TTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCTTCCCAATGCTGTCAAGCTCAGT 600

Qy     610 AGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTTAAGAAGGTAATCAGATTCCA 669
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 AGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTTAAGAAGGTAATCAGATTCCA 660

Qy     670 GGCTTCCTGATCGGAAAAGATTGGGGCAGGAACATTAAAGAGCCTTCATTATGACTGA 729
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GGCTTCCTGATCGGAAAAGATTGGGGCAGGAACATTAAAGAGCCTTCATTATGACTGA 720

Qy     730 ATTACACTCATCCTT 744
           |||||||||||||||
Db     721 ATTACACTCATCCTT 735
```

```
RESULT 5
US-10-210-130-127
; Sequence 127, Application US/10210130
; Publication No. US2004014053A1
GENERAL INFORMATION:
;   APPLICANT: Zerhusen, Bryan D.
;   APPLICANT: Patturajan, Meera
;   APPLICANT: Kekuda, Ramesh
;   APPLICANT: Miller, Charles B.
;   APPLICANT: Rieger, Daniel K.
;   APPLICANT: Pena, Carol E.A.
;   APPLICANT: Shimkets, Richard A.
;   APPLICANT: Li, Li
;   APPLICANT: Berghs, Constance
;   APPLICANT: Zhong, Mei
;   APPLICANT: Casman, Stacie J.
;   APPLICANT: Voss, Edward Z.
;   APPLICANT: Boldog, Ferenc L.
;   APPLICANT: Padigaru, Muralidhara
;   APPLICANT: Smithson, Glennda
;   APPLICANT: Ji, Weizhen
;   APPLICANT: Gorman, Linda
;   APPLICANT: Vernet, Corine A.M.
;   APPLICANT: Leite, Mario W.
;   APPLICANT: Guo, Xiaojia Sasha
;   APPLICANT: Anderson, David W.
;   APPLICANT: Spytek, Kimberly A.
;   APPLICANT: Gerlach, Valerie
;   APPLICANT: Burgess, Catherine E.
;   APPLICANT: Khramtsov, Nikolai V.
;   APPLICANT: Ort, Tatiana
;   APPLICANT: Ellerman, Karen
;   APPLICANT: Rastelli, Luca
;   APPLICANT: Agee, Michele L.
;   APPLICANT: Chaudhuri, Amitabha
```

```
;    APPLICANT:  Chant, John S.
;    APPLICANT:  DiPippo, Vincent A.
;    APPLICANT:  Edinger, Shlomit R.
;    APPLICANT:  Eisen, Andrew J.
;    APPLICANT:  Gangolli, Esha A.
;    APPLICANT:  Giot, Loic
;    APPLICANT:  Ooi, Chean Eng
;    APPLICANT:  Rothenberg, Mark E.
;    APPLICANT:  Spaderna, Steven K.
;    APPLICANT:  Hjalt, Tord
;    APPLICANT:  Liu, Xiaohong
;    APPLICANT:  Taupier, Raymond J., Jr.
;    APPLICANT:  Catterton, Elina
;    APPLICANT:  Shenoy, Suresh G.
;    TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
;    FILE REFERENCE: 21402-416C (Cura-716 SMT)
;    CURRENT APPLICATION NUMBER: US/10/210,130
;    CURRENT FILING DATE:  2002-08-01
;    PRIOR APPLICATION NUMBER: 60/309,501
;    PRIOR FILING DATE:  2001-08-02
;    PRIOR APPLICATION NUMBER: 60/316,508
;    PRIOR FILING DATE:  2001-08-31
;    PRIOR APPLICATION NUMBER: 60/354,655
;    PRIOR FILING DATE:  2002-02-05
;    PRIOR APPLICATION NUMBER: 60/310,291
;    PRIOR FILING DATE:  2001-08-03
;    PRIOR APPLICATION NUMBER: 60/383,887
;    PRIOR FILING DATE:  2002-05-29
;    PRIOR APPLICATION NUMBER: 60/310,951
;    PRIOR FILING DATE:  2001-08-08
;    PRIOR APPLICATION NUMBER: 60/323,936
;    PRIOR FILING DATE:  2001-09-21
;    PRIOR APPLICATION NUMBER: 60/381,039
;    PRIOR FILING DATE:  2002-05-16
;    PRIOR APPLICATION NUMBER: 60/311,292
;    PRIOR FILING DATE:  2001-08-09
;    PRIOR APPLICATION NUMBER: 60/311,979
;    PRIOR FILING DATE:  2001-08-13
;    Remaining Prior Application data removed - See File Wrapper or PALM.
;    NUMBER OF SEQ ID NOS: 369
;    SOFTWARE: CuraSeqList version 0.1
US-10-210-130-127
SEQ ID NO 127
;    LENGTH: 739
;    TYPE: DNA
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: CDS
;    LOCATION: (10)..(717)
US-10-210-130-127

Query Match             43.2%;  Score 731.6;  DB 17;  Length 739;
Best Local Similarity   99.5%;  Pred. No. 4.1e-191;
Matches   734;  Conservative   0;  Mismatches   4;  Indels   0;  Gaps   0;

Qy         10 AGTGGCAAGATGGCCGTCCCTGGTGTTGGGAGGTACTCGGTGAAGGTACTCGGTGTTCAGTGTT    69
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGTGGCAAGATCGCGTCCCTGGTGTTGGGAGGTACTCGGTGAAGGTACTCGGTGTTCAGTGTT    60

Qy         70 GGGAAATCTTCGTTAGTCCATCTCCTATGCCAAATCAAGTGCTGGGAAATCCATCATCATGG   129
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGGAAATCTTCGTTAGTCCATCTCCTATGCCAAATCAAGTGCTGGGAAATCCATCATCATGG   120

Qy        130 ACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATGTTCAGAGTTCATGATGTTCAGAGTTCATGATG   189
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATGTTCATGATGTTCAGAGTTCATGATG   180

Qy        190 AAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCGTGGGCGTGCCAGCAGCGTG   249
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCGTGGGCGTGCCAGCAGCGTG   240

Qy        250 AAAAGCACAAGAGCAGTATTCTCACAACTCCGTAAATGGTATTATTTTCGTACACGACTTA   309
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAAGCACAAGAGCAGTATTCTCACAACTCCGTAAATGGTATTATTTTCGTACACGACTTA   300
```

```
Qy        310 ACAAATAAGAAGTCCTCCCAAAACTTGGTCGTCGTTGGTCATTGGAAGCTCTCAACAGGGAT   369
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACAAATAAGAAGTCCTCCCAAAACTTGGTCGTCGTTGGTCATTGGAAGCTCTCAACAGGGAT   360

Qy        370 TTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTGTTGCTGAT   429
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGAT   420

Qy        430 AACCAAATACCACCTGTTGGTAATAGGGACTAAACTGACCAGATTCATGAAACAAAGCGC   489
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AACCAAATACCACCTGTTGGTAATAGGGACTAAACTGACCAGATTCATGAAACAAAGCGC   480

Qy        490 CATGAAGTTTTAACTAGGACTGCTTTCCTGGCTGCTGAGGATTTCAATCCAGAGAAATTAAT   549
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CATGAAGTTTTAACTAGGACTGCTTTCCTGGCTGCTGAGGATTTCAATCCAGAGAAATTAAT   540

Qy        550 TTGGACTGCACAAATCCACGTACTTAGCTGCAGGTTCTTCCAATGCTGTCGAAGCTCAGT   609
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TTGGACTGCACAAATCCACGTACTTAGCTGCAGGTTCCTCCAATGCTGTCGAAGCTCAGT   600

Qy        610 AGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTAAGAGAAGTAATCAGATTCCA   669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTAAGAGAAGTAATCAGATTCCA   660

Qy        670 GGCTTCCTGATCGGAAAAGATTGGGGCAGGAACATTAAAGAGCCTTCATTATGACTGA   729
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGCTTTCCTGATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGCCCTTCATTATGACTGA   720

Qy        730 ATTACACTCATCCTTTGG   747
              ||||||||||||||||||
Db        721 ATTACACTCATCCTAAGG   738
```

```
RESULT 6
US-09-945-173-3
;    Sequence 3, Application US/09945173
;    Patent No. US2002012568A1
;    GENERAL INFORMATION:
;    APPLICANT: Millennium Pharmaceuticals, Inc.
;    APPLICANT: Meyers, Rachel
;    TITLE OF INVENTION: 47324, A NOVEL HUMAN G-PROTEIN AND USES
;    TITLE OF INVENTION: THEREFOR
;    FILE REFERENCE: 38155-20035.00
;    CURRENT APPLICATION NUMBER: US/09/945,173
;    CURRENT FILING DATE: 2001-08-31
;    PRIOR APPLICATION NUMBER: US 60/229,293
;    PRIOR FILING DATE: 2000-09-01
;    NUMBER OF SEQ ID NOS: 11
;    SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 3
;    LENGTH: 711
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-09-945-173-3

Query Match             42.0%;  Score 711;  DB 9;  Length 711;
Best Local Similarity  100.0%;  Pred. No. 2e-185;
Matches   711;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy         19 ATGGGCGTCCCTGGATCGGGTGAAGGTACTCGGTGTTGGGAGACTCAGTGTGTTGGGAAATCT    78
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGCGTCCCTGGATCGGGTGAAGGTACTCGGTGTTGGGAGACTCAGTGTGTTGGGAAATCT    60

Qy         79 TCGTTAGTCCATCTCCTATGCCAAATCAAGTGCTGGGAAATCCATCATCATGGACTGTGGGC   138
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCGTTAGTCCATCTCCTATGCCAAATCAAGTGCTGGGAAATCCATCATCATGGACTGTGGGC   120

Qy        139 TGCTCAGTGGATGTCAGAGTTCATGATTCAAGAAGAAGACTTAC   198
              ||||||||||||||||||||||||||||||||||||||||||||
Db        121 TGCTCAGTGGATGTCAGAGTTCATGATTCAAGAAGAAGACTTAC   180

Qy        199 TACATAGAATTATGGGATGTTGGAGGCTCGTGGGCAGTGCCAGCAGCGTGAAAAGCACA   258
```

```
Db  181 TACATAGAATTATGGATGGTCGAGGCTCTGTGGGCAGTGCCAGCAGCGCGTGAAAAGCACA 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  128 GGACTGTGGGCTGCTCAGTGACTCAGAGTTCATGATTACAAAGAGGAACCCCAGAAG 187
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 GGACTGTGGGCTGCTCAGTGACTCAGAGTTCATGATTACAAAGAGGAACCCCAGAAG 180

Qy  259 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTAACAAATAAG 318
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  188 AGAAGACCTACTACATAGAAATTATGGGACGTCTGTGGGCAGTGCCAGCAGCG 247
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTAACAAATAAG 240
Db  188 AGAAGACCTACTACATAGAAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCG 240

Qy  248 TGAAAAGCACAAGACAGTATTCTACAACTCCGTAAATGTGTATTATTTTCGTACACGACT 307
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  241 TGAAAAGCACAAGACAGTATTCTACAACTCCGTAAATGTGTATTATTTTCGTACACGACT 300

Qy  308 TAACAAATAAGAAGTCCTCCCAAAACTCATGAAACAAAGCGCCATGAAGTT 367
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  301 TAACAAATAAGAAGTCCTCCCAAAACTCTCCCAAAACTCAACAGGG 360

Qy  368 ATTTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTGTGACAAATGGGATTTGCTG 427
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  361 ATTTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGGACAAATGGGATTTGCTG 420

Qy  428 ATAACCAAATACCACCTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGC 487
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  421 ATAACCAAATACCACCTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGC 480

Qy  488 GCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTCGTTCAATCCAGAGAAATTA 547
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  481 GCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTCGTTCAATCCAGAGAAATTA 540

Qy  548 ATTTGGACTGCACAAATCCACGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCA 607
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  541 ATTTGGACTGCACAAATCCACGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAA-CTCA 599

Qy  608 GTAGGTTTTTTGAT-AGGGCATAGAAGAGAGATAC-TTTAAGAAGTAATCNGATCC 657
            :||||||||||||| |||| |||||||||||||||||||||||||||||||||||
Db  600 RTAGGTTTTTTGAT-AGGGCATAGAAGAGAGATAC-TTTAAGAAGTAATCNGATCC 657

Qy  668 CAGGCTTTCCTGATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGCCT 716
            |||||        ||||||||||||||||||||||||||||||||||||
Db  658 GAGGCTCTTGACTCG---AAAAGATTGGGGCCGGACCATTAAAGCGCTT 703
```

RESULT 7
US-09-764-868-321
; Sequence 321, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ32
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 321
; LENGTH: 959
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (652)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (705)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-321

Query Match   37.1%; Score 628.4; DB 9; Length 959;
Best Local Similarity 96.5%; Pred. No. 1.6e-162;
Matches 684; Conservative 1; Mismatches 18; Indels 6; Gaps 4;

```
Qy    8 GCAGTGCCAAGATGGCCGTCCCTGGATCGGTCGAAGGTACTCGGTGTTGGGAGACTCAGGTG 67
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GCAGTGCCAAGATGGCCGTCCCCTGGATCGCGGTGGAAGGTACTCGGTGTTGGGAGACTCAGGTG 60

Qy   68 TTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCAT 127
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCAT 120
```

RESULT 8
US-09-918-995-23516
; Sequence 23516, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; SOFTWARE: FastSEQ for Windows Version 3.0
; NUMBER OF SEQ ID NOS: 38054
; SEQ ID NO 23516
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(504)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-23516

Query Match   26.2%; Score 444; DB 10; Length 504;
Best Local Similarity 96.4%; Pred. No. 9.7e-112;
Matches 450; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

```
Qy   20 TGGCGTCCCTCGGATCGGTCGAGGTGAGGTACGTCAGGTGTTGGGAGACTCAGGTGTTGGGAAATCTT 79
            ||   ||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Qy   37 TTGGNANNNNGGCTCGGNNGAAGGACTCAGCTTGGGAGACTCAGGTGTTGGGAAATCTT 96
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   80 CGTTAGTCCATCTCCTATGCCAAAATCAGTCGTCGGAAATCCATCATGGACTGTGGGCT 139

Qy   97 TTGGGAAATCTT 156

RESULT 9
US-09-983-965-2369
; Sequence 2369, Application US/09983965
; Patent No. US2002013716OA1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 2369
; LENGTH: 387
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 64-LIB3057-021-Q1-K1-H8
US-09-983-965-2369

Query Match   14.9%; Score 252.8; DB 9; Length 387;
Best Local Similarity   84.5%; Pred. No. 5.6e-59;
Matches 333; Conservative   0; Mismatches 52; Indels 9; Gaps 4;

Qy   760 GCAGTGGCAGTTTTCACAGCTCATCTTGCTGTGT-TCAATTATTACCATCACAGCCTTT 818
          ||||| |||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   1 GCGGTGGCAGTTTTTCACAGCTTCTTCTTGTGTCGATTATTAATGTCTCAGCCTTT 60

Qy   819 TAACAAATCATC-TTAAAATGCTACCCTTCAGCCTTAATGGAAAAATGAAAG 877
          ||| |||||||| ||||||||||||||||||||||| ||||||| |||||
Db   61 TAATGAAATCATCTTTAAAATGCCACCCTTCAGCCTTTAGTGGAAAACTGAAAG 120

Qy   878 GAAGTGACAATACGGGAGAGGTCCAAACTTTGTCCCTGTTCTCGTGTTCCTTACCTTTCTG 937
          ||||| |||||||||||||||||||||||||||||||||||||||| ||| |||||| |||
Db   121 GAAGTAACAGCGTGGGAGAGCTCCAGACTTGGTCC----TCATTCTATTTTTCACCCTTCTG 176

Qy   938 TCCCTGTGTATAGATTATGTAAAGCCTTGTGTGTAAATATGAGATGTTGTCAAAATGATGC 997
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177 TCCCTGTTTATAGATTATGTAAAGCCTTGTGTGTAAATATGAGA---TGTCAAAATGATGC 233

Qy   998 AGTAAATGAGCAATGACAGTGTACTGCAGAGAAAATTTACTCTTGCTAGAACTCGGAGGG 1057
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   234 AGTAAATGAGCAGTGACAGAATGCTGCAGAGAAAATTTACTCTTGCCTAGAACTCGGAGGA 293

Qy 1058 TTTTTATGGGTCTGTAATTTCCCACACTCATTGCTGAAAGCTTAATTAAGTACTTCAAA 1117
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   294 TTTTTATGGGTCTGTAATTTCCCACACTCATTGCTGAAAGCTTAATTAAGTACTTCAGA 353

Qy 1118 AACGTATCTCCATTGTTTTACCTTCTGAGGGGA 1151
          ||||||||||||||||||||||||||||||||
Db   354 AATGTATCTCCATTGTTTTACCTTCTGAGGGGAA 387

RESULT 10
US-09-908-975-10523
; Sequence 10523, Application US/09908975
; Publication No. US2003016583A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, Avi
; APPLICANT: WASSERMAN, Alon
; APPLICANT: MINTZ, Eli
; APPLICANT: MINTZ, Liat
; APPLICANT: FAIGLER, Simchon
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE
; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10523
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-908-975-10523

Query Match   3.5%; Score 60; DB 10; Length 60;
Best Local Similarity   100.0%; Pred. No. 3.6e-06;
Matches 60; Conservative   0; Mismatches 0; Indels 0; Gaps 0;

Qy   891 GGGAGGTCCAAACTTTGTCCCTGTTCTCGTGTTCCTTACCTTTCTGTCCCTGTGTATAG 950
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGGAGGTCCAAACTTTGTCCCTGTTCTCGTGTTCCTTACCTTTCTGTCCCTGTGTATAG 60

RESULT 11
US-10-425-115-2369
; Sequence 2369, Application US/10425115
; Publication No. US2004021427242A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 2369
; LENGTH: 1626
; TYPE: DNA
; ORGANISM: Zea mays

; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_10216lC.1
US-10-425-115-2369

Query Match 3.2%; Score 54.6; DB 20; Length 1626;
Best Local Similarity 60.4%; Pred. No. 0.00074;
Matches 90; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy    21 GGCGTCCCTGGATCGGGTGAAGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTC 80
         || ||| | |||  ||||   || ||| ||||||||||||||||||| ||||||
Db   342 GCCGCCCTGCGGCCAGGTTCGTGTACTCGTCGTTGGCGACTCAGGTGTTGGCAAATCTTC 401

Qy    81 GTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATCATGGACTGTGGGCTG 140
         ||||   || |  ||||| || ||| |||||| || |||   |||||| |||||
Db   402 ATTGGTGCATCTCCTTACTGAAAGATTCTGCAGTGGCTCGACCAGCCCAAACAATTGGGTG 461

Qy   141 CTCAGTGGATGTCAGAGTTCATGATTACA 169
         |||| |||||| |||||||||||||||
Db   462 TGCAGTTGGTGTTAAACATATTACTTACA 490
```

RESULT 12
US-10-767-701-25416
; Sequence 25416, Application US/10767701
; Publication No. US200401726B4A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 25416
; LENGTH: 664
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 30951803
US-10-767-701-25416

Query Match 3.0%; Score 51.4; DB 19; Length 664;
Best Local Similarity 59.1%; Pred. No. 0.0034;
Matches 88; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

```
Qy    21 GGCGTCCCTGGATCGGGTGAAGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTC 80
         || ||| | |||  ||||   || ||| ||||||||||||||||||| ||||||
Db   398 GCCGCCCTGCGGCCAGGTGCGTGTCGTCGTTGGCGACTCAGGGGTTGGTAAATCTTC 457

Qy    81 GTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATCATGGACTGTGGGCTG 140
         ||||   || | ||||| || ||| |||||| || |||   |||||| |||||
Db   458 ATTGGTGCATCTCCTTATTGAAAGGTTCTGCAGTGGCTCGACCAGCCCAAACAATTGGGTG 517

Qy   141 CTCAGTGGATGTCAGAGTTCATGATTACA 169
         |||| ||||| |||||||||||||||
Db   518 TGCAGTTGGTGTTAAACATATCACTTACA 546
```

RESULT 13
US-10-425-114-26886
; Sequence 26886, Application US/10425114
; Publication No. US200400348B8A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 26886
; LENGTH: 1643
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4666-008-F5_FLI
US-10-425-114-26886

Query Match 3.0%; Score 50.2; DB 18; Length 1643;
Best Local Similarity 59.4%; Pred. No. 0.012;
Matches 85; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```
Qy    26 CCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTTAG 85
         |||| || | |||| ||| |||  || ||||||||||||||||||||
Db   325 CCTTCGGCCAAGTGCGCGTCCTCATCGTCGGTGATTCAGGTGTTGGGAAATCTTCATTGG 384

Qy    86 TCCATCCTATGCCAAAATCAAGTGCTGGGAAATCATCATGCCATTGCTGCCATCTCAAACAGTAGGATGCACTGCTCAG 145
         ||||| ||||||||||||||||| | ||| ||||||||||||||||||||||||||
Db   385 TGCATCTCATTTTGAAAAGTTCTGCCATTGCTGCCATCTCAAACAGTAGGATGCACTG 444

Qy   146 TGGATGTCAGAGTTCATGATTAC 168
         ||| |||||||||||||||
Db   445 TGGGCATTAAACATGTTACTTAC 467
```

RESULT 14
US-10-425-115-75255
; Sequence 75255, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 75255
; LENGTH: 1643
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_168664C.1
US-10-425-115-75255

Query Match 3.0%; Score 50.2; DB 20; Length 1643;
Best Local Similarity 59.4%; Pred. No. 0.012;
Matches 85; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```
Qy    26 CCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTTAG 85
         |||| || | |||| ||| |||  || ||||||||||||||||||||
Db   325 CCTTCGGCCAAGTGCGCGTCCTCATCGTCGGTGATTCAGGTGTTGGGAAATCTTCATTGG 384

Qy    86 TCCATCCTATGCCAAAATCAAGTGCTGGGAAATCATCATGCCATTGCTGCCATCTCAAACAGTAGGATGCACTGCTCAG 145
         ||||| ||||||||||||||||| | ||| ||||||||||||||||||||||||||
Db   385 TGCATCTCATTTTGAAAAGTTCTGCCATTGCTGCCATCTCAAACAGTAGGATGCACTG 444

Qy   146 TGGATGTCAGAGTTCATGATTAC 168
         ||| |||||||||||||||
Db   445 TGGGCATTAAACATGTTACTTAC 467
```

RESULT 15
US-10-437-963-25042/c
; Sequence 25042, Application US/10437963
; Publication No. US20040123343A1

; GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
;   APPLICANT: Zhou, Yihua
;   APPLICANT: Cao, Yongwei
;   APPLICANT: Wu, Wei
;   APPLICANT: Boukharov, Andrey A.
;   APPLICANT: Barbazuk, Brad
;   APPLICANT: Li, Ping
;   TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
;   TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53221)B
;   CURRENT APPLICATION NUMBER: US/10/437,963
;   CURRENT FILING DATE: 2003-05-14
;   NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 25042
;   LENGTH: 1584
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_29969C.1
US-10-437-963-25042

Query Match   2.7%;   Score 46.2; DB 19;   Length 1584;
Best Local Similarity 54.4%;   Pred. No. 0.15;
Matches   93;   Conservative   0;   Mismatches   78;   Indels   0;   Gaps   0;

Qy   2 GGGAAGGCAGTGGCAAGATGGCGTCCCTGGATCGGCGGTGAAGGTACTGGTGTTGGGAGACT 61
            |||| || ||| | | | |||| ||| ||| ||||||
Db   1255 GGGACCTCAACGGGCGGCGGCACCCCCTGCCGCCAGGTGCCGGTGCGTGCTCGTCGGCGACT 1196

Qy   62 CAGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAAATCAAGTGCTGGGAAATC 121
            ||||||||| |||||||| || || | |||||| | | || | | |
Db   1195 CAGGTGTGGGGAAATCATCATCATTGGTTCATCTCATTCGAAAGGCTCTGCAATTGCTCGAC 1136

Qy   122 CATCATGGACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAG 172
            ||| | || | | |||| |||| |||| ||||| ||
Db   1135 CACCCCAAACGATTGGATGCCGCAGTTGATGTTAAATATATTTTATTGAAG 1085

Search completed: June 30, 2005, 23:57:27
Job time : 1188.77 secs

```
                GenCore version 5.1.6
        Copyright (c) 1993 - 2005  Compugen Ltd.

OM protein - protein search, using sw model

Run on:    June 30, 2005, 20:25:54 ; Search time 24 Seconds
                            (without alignments)
                            946.132 Million cell updates/sec

Title:      US-09-945-173-2
Perfect score: 1231
Sequence:   1 MASLDRVKVLVLGDSGVGKS.........PGFPDRKRFGAGTLKSLHYD 236
Scoring table: BLOSUM62
            Gapop 10.0 , Gapext 0.5

Searched:   283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters:     283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   PIR 79:*
             1: .pir1:*
             2: .pir2:*
             3: .pir3:*
             4: .pir4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	259	21.0	416	2	T31486	hypothetical prote
2	171.5	13.9	217	2	S36365	GTP-binding protei
3	169	13.7	203	2	S51495	GTP-binding protei
4	167	13.6	203	2	B38202	GTP-binding protei
5	166.5	13.5	202	2	S41430	GTP-binding protei
6	164.5	13.4	201	2	JC2488	GTP-binding protei
7	164.5	13.4	206	2	JC4107	membrane vesicle t
8	163.5	13.3	196	2	PS0279	GTP-binding protei
9	163.5	13.3	258	2	B86153	ARA-5 [imported] -
10	162.5	13.2	208	2	A34716	GTP-binding protei
11	162	13.2	221	2	H71444	GTP-binding protei
12	162	13.2	234	2	S38083	probable purine nu
13	161	13.1	203	2	B34716	GTP-binding protei
14	160.5	13.0	203	2	T50323	ypt1-related prote
15	160.5	13.0	206	2	S04590	GTP-binding protei
16	160.5	13.0	209	2	T31551	hypothetical prote
17	160	13.0	205	2	S36368	GTP-binding protei
18	159	12.9	205	2	T03629	GTP-binding protei
19	159	12.9	206	2	S33531	GTP-binding protei
20	158.5	12.9	203	2	S34253	GTP-binding protei
21	158.5	12.9	208	2	A44334	GTP-binding protei
22	158	12.8	201	2	JC5337	GTP-binding protei
23	158	12.8	202	2	S38740	GTP-binding protei
24	157.5	12.8	201	2	S06147	GTP-binding protei
25	157.5	12.8	201	2	S39565	GTP-binding protei
26	157	12.8	206	2	I78851	GTP-binding protei
27	157	12.8	206	2	C96529	probable RAB7 GTP-
28	156	12.7	201	2	D38625	GTP-binding protei
29	156	12.7	206	2	H96562	hypothetical prote
30	155.5	12.6	200	2	S12790	GTP-binding protei
31	155.5	12.6	208	2	S40207	GTP-binding protei
32	155.5	12.6	208	2	A38202	Ras-related protei
33	155.5	12.6	218	2	T31656	ras-related protei
34	155	12.6	205	2	T40425	GTP-binding protei
35	155	12.6	205	2	S38339	GTP-binding protei
36	155	12.6	211	2	S34729	GTP-binding protei
37	155	12.6	216	1	JC4108	GTP-binding protei
38	154.5	12.6	205	1	TVDGYP	GTP-binding protei
39	154.5	12.6	205	1	TVHUYP	GTP-binding protei
40	154.5	12.6	205	2	TVMSYP	GTP-binding protei
41	154.5	12.6	217	2	S30273	GTP-binding protei
42	154	12.5	211	1	T29035	hypothetical prote
43	154	12.5	215	1	TVBYQ4	GTP-binding protei
44	154	12.5	215	2	T14565	GTP-binding protei
45	154	12.5	215	2	S57462	GTP-binding protei

ALIGNMENTS

RESULT 1
T31486
hypothetical protein Y116A8C.10 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C;Accession: T31486
R;McMurray, A.
submitted to the EMBL Data Library, October 1999
A;Reference number: Z21041
A;Accession: T31486
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-416 <WIL>
A;Cross-references: UNIPROT:Q9U2V7; EMBL:AL117204; PIDN:CAB55120.1; CESP:Y116A8C.10
C;Genetics:
A;Gene: CESP:Y116A8C.10
A;Experimental source: clone Y116A8C
A;Introns: 25/2; 63/3; 103/2; 125/1

```
Query Match        21.0%;   Score 259;  DB 2;   Length 416;
Best Local Similarity  32.2%;  Pred. No. 3.8e-14;
Matches 76;  Conservative 32;  Mismatches 86;  Indels 42;  Gaps 10;

Qy    1 MASLDR--VKVLVLGDSGVGKSSLVHLLCQNQVLGNP------SWTVGCSVDVRVHDYK 51
          :|||| :|||||||||||||:| : :::             :|:|:|:|:
Db  162 MSSLDESSTKILVLGDSCVGKTSLCHCIAGG---GEPPGGGRSFDSTIGATVVMAWHYR 218

Qy   52 EGTPBRKTYYIELWDIGGVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRW-- 109
         |||||:|:||: ||||:|:|||          :        ||:||:
Db  219 AGTPEQRTELLELWDIGMVAHRQAAQ---VPFEGAVGAILVHDLTNKRSEENLATWLT 274

Qy  110 SLRALNRDLVPTGVLVTNGDYDQEQFADNQIPLLVIGTKLDQI-HE----TKRHEVLIRT 164
         :|:         ||::||||   :            :|:|:|:|:|   ||    :|::|||:|
Db  275 MLDGKPRGAAPKSSKDPAAVALKVDIBSCNIPVLIVGTKADLVPHKGPVSYDRYDGLGNA 334

Qy  165 AFLAEDFNPEEINLDCTNPRYLAAGSSNAVKLSRFFDKVIEK------RYFLREG 213
          :           :|:     ||:       |||||
Db  335 --------NSAISVWKPRNRRKNSKKPMKIGR--KLVVEKEKVKKFGRQYLRRG 378
```

RESULT 2
S36365
GTP-binding protein yptV2 - Volvox carteri
C;Species: Volvox carteri
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 09-Jul-2004
C;Accession: S36365
R;Fabry, S.; Jacobsen, A.; Huber, H.; Palme, K.; Schmitt, R.
Curr. Genet. 24, 229-240, 1993
A;Title: Structure, expression, and phylogenetic relationships of a family of ypt genes
A;Reference number: S36365; MUID:94037148; PMID:8221932
A;Accession: S36365
A;Molecule type: DNA

A;Residues: 1-217 <FAB>
A;Cross-references: UNIPROT:P36861; GB:L08128; NID:g409161; PIDN:AAA34251.1; PID:g409162
A;Gene: yptV2
A;Introns: 23/1; 47/2; 81/1; 121/3; 166/3
C;Superfamily: ras transforming protein; translation elongation factor Tu homology
C;Keywords: GTP binding; nucleotide binding; P-loop
F;14-129/Domain: translation elongation factor Tu homology <ETU>
F;20-27/Region: nucleotide-binding motif A (P-loop)
F;66-73/Region: GTP binding #status predicted
F;123-130/Region: GTP binding #status predicted
F;153-159/Region: GTP binding #status predicted

Query Match   13.9%; Score 171.5; DB 2; Length 217;
Best Local Similarity   27.3%; Pred. No. 4e-07;
Matches 45; Conservative 36; Mismatches 55; Indels 29; Gaps   6;

Qy   7 VKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTYYIEBLWD 66
               :|:||||||||||||:|- | :|:::|| : :::||
Db   14 IKLLLVGDSGVGKSCLLLRFTDDMFTSSFITTIG--IDFKI---KKVDVDGKLVKLQIW 68

Qy   67 VGGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLVT 126
              || |||| ||| |||| ::|:::|| || :||
Db   69 TAGQ---ERFRTTTSAYYRGAQGIILVDITDEASFNNVRNWM-----RNI-------- 111

Qy   127 NGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDF 171
                |||| : |:| | :||||| :|| || ||
Db   112 ___EQHASDVNKILVGNKLD-LAEDKRVVSARGQALADEF 149

RESULT 3
S51495
GTP-binding protein RYL1 - yeast (Yarrowia lipolytica)
N;Alternate names: SEC4 protein homolog; small Rab-related protein
C;Species: Yarrowia lipolytica, Candida lipolytica
C;Date: 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change 09-Jul-2004
C;Accession: S51495; S70850
R;Pertuiset, B.; Beckerich, J.M.; Gaillardin, C.
A;Title: Molecular cloning of Rab-related genes in the yeast Yarrowia lipolytica. Analyse
Curr. Genet. 27, 123-130, 1995
A;Reference number: S51495; MUID:95308533; PMID:7788714
A;Accession: S51495
A;Molecule type: DNA
A;Residues: 1-203 <PER>
A;Cross-references: UNIPROT:P41924; EMBL:L06969;
R;Pertuiset, B.; Beckerich, J.M.; Gaillardin, C.
submitted to the EMBL Data Library, December 1992
A;Description: RYL1 gene a SEC4 homolog in the yeast Yarrowia lipolytica.
A;Reference number: S70850
A;Accession: S70850
A;Molecule type: DNA
A;Residues: 1-110,'S',112-203 <PEW>
A;Cross-references: EMBL:L06969; NID:g173256; PIDN:AAA35245.1; PID:g173257
A;Gene: RYL1
C;Superfamily: ras transforming protein; translation elongation factor Tu homology
C;Keywords: GTP binding; lipoprotein; membrane trafficking; nucleotide binding; P-loop;
F;18-25/Region: translation elongation factor Tu homology <ETU>
F;124-127/Region: GTP-binding NKXD motif
F;202,203/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match   13.7%; Score 169; DB 2; Length 203;
Best Local Similarity   24.2%; Pred. No. 5.9e-07;
Matches 48; Conservative 35; Mismatches 73; Indels 42; Gaps   7;

Qy   7 VKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTYYIELWD 66
               :|:||||||||| |: | : :: | : || :: :|
Db   12 IKLLLGDSGVGKSCLLLRFCEDQF--TPSFITTIGIDFKIRTDIGNQRVK---LQVWD 66

Qy   67 VGGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLVT 126
              || |||| ||| |||| ::|:::|| || :||
Db   67 TAGQ---ERFRTTTAYYRGAMGILLVYDVTDEKSFNNIENW------------ 105

Qy   127 NGDYDQEQ-FADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNPEBINLDCTNPRY 185
               ||||| : |:| | :||||| :|| || || :: ||
Db   106 ---YQNVQTYANEGVBLILVGNKCDL--DEKRVVSTBQGQALADKFGIP---------F 150

Qy   186 LAAGSSNAVKLSRFFDKV 203
              |:: |||| ||||||
Db   151 LEASSKTNINVEECFYSV 168

RESULT 4
B38202
GTP-binding protein - maize
C;Species: Zea mays (maize)
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: B38202
R;Palme, K.; Diefenthal, T.; Vingron, M.; Sander, C.; Schell, J.
Proc. Natl. Acad. Sci. U.S.A. 89, 787-791, 1992
A;Title: Molecular cloning and structural analysis of genes from Zea mays (L.) coding for
A;Reference number: A38202; MUID:92115746; PMID:1731354
A;Accession: B38202
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-203 <PAL>
A;Cross-references: UNIPROT:Q05737; GB:X63278; NID:g287834; PIDN:CAA44919.1; PID:g287835
A;Note: sequence extracted from NCBI backbone (NCBIN:76123, NCBIP:76124)
C;Superfamily: ras transforming protein; translation elongation factor Tu homology
C;Keywords: GTP binding; membrane trafficking; nucleotide binding; P-loop
F;9-124/Domain: translation elongation factor Tu homology A (P-loop) <ETU>
F;15-22/Region: nucleotide-binding motif A (P-loop)
F;121-124/Region: GTP-binding NKXD motif
F;151-153/Region: GTP-binding SAK/L motif
F;21,22,40,121,122,124,151/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #stat

Query Match   13.6%; Score 167; DB 2; Length 203;
Best Local Similarity   25.6%; Pred. No. 8.7e-07;
Matches 50; Conservative 40; Mismatches 61; Indels 44; Gaps   8;

Qy   8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEE--KTYYIELW 65
              |:||||||||||: | : ::| : ||-----DFKIRTVEQDGKTIKLQIW 62
Db   10 KLLLIGDSGVGKSCLLRFADDSYLDSYISTIGV------DFKIRTVEQDGKTIKLQIW 62

Qy   66 DVGGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLV 125
              | || |||| ||| |||| ::|:::|| || :||
Db   63 DTAGQ---ERFRTTTSSYYRGAHGIIIVDVTDQESFNNVKQW------LN------- 104

Qy   126 TNGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNPEBINLDCTNPRY 185
               |||||: |:| | :||||| :|| || || :|:|
Db   105 ----EIDRYASDNVNKLLVGNKSDL--TANKVVATETAKAFAD---EMGIP-----F 147

Qy   186 LAAGSSNAVKLSRFF 200
              |:: |||| ||||
Db   148 METSAKNATNVQQAF 162

RESULT 5
S41430
GTP-binding protein, ras-like (clone vfa-ypt1) - fava bean
N;Alternate names: guanine nucleotide regulatory protein
C;Species: Vicia faba (fava bean)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C;Accession: S41430
R;Saalbach, G.; Thielmann, J.
submitted to the EMBL Data Library, January 1994
A;Description: Sequences of cDNA clones from cotyledons of Vicia faba encoding ypt/rab-re
A;Reference number: S41430
A;Accession: S41430
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-202 <SAA>
A;Cross-references: UNIPROT:Q08155; EMBL:Z29590; NID:g452358; PIDN:CAA82707.1; PID:g45235
A;Experimental source: clone vfa-ypt1
C;Superfamily: ras transforming protein; translation elongation factor Tu homology

C;Keywords: GTP binding; lipoprotein; membrane trafficking; nucleotide binding; P-loop
F;9-124/Domain: translation elongation factor Tu homology <ETU>
F;15-22/Region: nucleotide-binding motif A (P-loop)
F;121-124/Region: GTP-binding NKXD motif
F;151-153/Region: GTP-binding SAK/L motif
F;21,22,40,121,122,124,151/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta

Query Match   13.5%; Score 166.5; DB 2; Length 202;
Best Local Similarity   26.8%; Pred. No. 9.5e-07;
Matches   51; Conservative   37; Mismatches   65; Indels   37; Gaps   7;

Qy   8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEE--KTYYIELW 65
            |:|:|||||||||| : : : ::|| || ::|| || || || :::||
Db   10 KLLIIGDSGVGKSCLLLRPADDSYIDSYISTIGV------DFKIRTVEQDGKTIKLQIW 62

Qy   66 DVGGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLV 125
            |||| |||||||:|:||:||| :|||| :: :|:|:|
Db   63 DTAGQ---ERFRTITSSYYRGAHGIIVYDVTDEESFNNVKQWLSEI------------- 106

Qy   126 TNGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLABDFNP--EEINLDCTN- 182
                       ::   :::| ||:|:|| |::|
Db   107 ------DRYASDNVNKLLVGNKCD-LTENRAVPYETAKAFADEIGIPFMETSAKDSTNV 158

Qy   183 -PRYLAAGSS 191
            ::|
Db   159 EQAFMAMASS 168

RESULT 6
JC2488
GTP-binding protein H-ray - human
C;Species: Homo sapiens (man)
C;Date: 16-Mar-1995 #sequence_revision 26-May-1995 #text_change 09-Jul-2004
R;Zhu, A.X.; Zhao, Y.; Flier, J.F.
Biochem. Biophys. Res. Commun. 205, 1875-1882, 1994
A;Title: Molecular cloning of two small GTP-binding proteins from human skeletal muscle.
A;Reference number: JC2487; MUID:95110337; PMID:7811277
A;Accession: JC2488
A;Molecule type: mRNA
A;Residues: 1-201 <ZHU>
A;Cross-references: UNIPROT:Q15286; EMBL:X79781; NID:g763121; PIDN:CAA56177.1; PID:g7631
A;Experimental source: skeletal muscle
C;Genetics:
A;Gene: GDB:SMGTP2
A;Cross-references: GDB:698171
A;Superfamily: ras transforming protein; translation elongation factor Tu homology
C;Keywords: GTP binding; lipoprotein; nucleotide binding; P-loop; prenylated cysteine
F;9-123/Domain: translation elongation factor Tu homology <ETU>
F;15-22/Region: nucleotide-binding motif A (P-loop)
F;120-123/Region: GTP-binding NKXD motif
F;150-152/Region: GTP-binding SAK/L motif
F;21,22,40,120,121,123,150/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #status predicted
F;200,201/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match   13.4%; Score 164.5; DB 2; Length 201;
Best Local Similarity   25.9%; Pred. No. 1.4e-06;
Matches   53; Conservative   36; Mismatches   63; Indels   53; Gaps   9;

Qy   8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEE--EKTYYIEL 64
            |:|:||||||||||| : : ::::|| || ::|| || || || :::
Db   10 KLLIIGDSGVGKSSLLLRFADNTPSGSYITTIGV------DFKIRTVEINGEKV-KLQI 61

Qy   65 WDVGGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVL 124
            :|||| |||||||:|:||:||| :|||| :: :|:|:|
Db   62 WDTAGQ---ERFRTITSTYYRGTHGIVVVDVTSAESFVNVKRW-LHEINQNC------- 110

Qy   125 VTNGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAF-------------LAED 170
                        ::   :::| ||:|:|| |::|
Db   111 ------DDVCRILVGNKND--DPERKVVETEDAYKFAGQMGIQLFETSAKEN 154

Qy   171 FNPEEINLDCTNPRYLAAGSSNAVK 195
             :: ||||| | |: | |||
Db   155 VNVEEM-FNCITELVLRAKKDNLAK 178

RESULT 7
JC4107
membrane vesicle transport protein ypt C5 - Chlamydomonas reinhardtii
N;Alternate names: ras-like yptC5 protein; small G protein yptC5
C;Species: Chlamydomonas reinhardtii
C;Date: 21-Nov-1995 #sequence_revision 08-Feb-1996 #text_change 09-Jul-2004
C;Accession: JC4107
R;Dietmaier, W.; Fabry, S.; Huber, H.; Schmitt, R.
Gene 158, 41-50, 1995
A;Title: Analysis of a family of ypt genes and their products from Chlamydomonas reinhar
A;Reference number: JC4105; MUID:95309723; PMID:7789809
A;Accession: JC4107
A;Molecule type: mRNA
A;Residues: 1-206 <DIE>
A;Cross-references: UNIPROT:Q39573; GB:U13170; NID:g806725; PIDN:AAA82728.1; PID:g806726
C;Comment: This protein plays an essential role in the regulation of intracellular membr
C;Genetics:
A;Gene: yptC5
A;Introns: 18/1; 60/3; 109/3; 136/3; 185/1
C;Superfamily: ras transforming protein; translation elongation factor Tu homology
C;Keywords: GTP binding; nucleotide binding; P-loop
F;9-128/Domain: translation elongation factor Tu homology <ETU>
F;15-22/Region: nucleotide-binding motif A (P-loop)
F;37-45/Domain: effector #status predicted <EFF>
F;125-128/Region: GTP-binding NKXD motif
F;158-160/Region: GTP-binding SAK/L motif

Query Match   13.4%; Score 164.5; DB 2; Length 206;
Best Local Similarity   24.2%; Pred. No. 1.4e-06;
Matches   55; Conservative   35; Mismatches   62; Indels   75; Gaps   9;

Qy   7 VKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDY--KEGTPEEKTYYIEL 64
            :||:|||||||||||| : : : ::|| || ::||
Db   9 LKVIILGDSGVGKTSLMNQYVQKFTKEYKATIGA--------DFLTKIEVDDKKVTMQI 61

Qy   65 WDVGG----SVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLV 119
            ||||| :||| |:|:||: | :||||
Db   62 WDTAGQERFQSLGSA-------FYRGADCCVLVFDVNNAKSFDDLDNWRDE-------- 105

Qy   120 PTGVLVTNGDYDQEQFADNQIPLLVIGTKLD---------QIHETK----------- 156
                                  ::   :::| ||
Db   106 ---FIIQAGPPDPDNF-----PFMVLGNKIDENGGSSRQVSEKKAKAWCASKGSIPYFET 157

Qy   157 --RHEVLTRTAFLA-------------EDFNPEEINLDCTNPRYLAAG 189
              : | :: :|
Db   158 SAKEDINVEAAFTCITRNALRNEKEEBLFMPDAVDMNTTATQRKRAG 204

RESULT 8
PS0279
GTP-binding protein ara-5 - Arabidopsis thaliana (fragment)
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 09-Jul-2004
C;Accession: PS0279; JQ0910
R;Anai, T.; Hasegawa, K.; Watanabe, Y.; Uchimiya, H.; Ishizaki, R.; Matsui, M.
Gene 108, 259-264, 1991
A;Title: Isolation and analysis of cDNAs encoding small GTP-binding proteins of Arabidop
A;Reference number: JS0639; MUID:92084144; PMID:1748311
A;Accession: PS0279
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-196 <ANA>
A;Cross-references: UNIPROT:P28188; GB:D01027
R;Anai, T.; Hasegawa, K.; Watanabe, Y.; Uchimiya, H.; Ishizaki, R.; Matsui, M.
submitted to JIPID, May 1991
A;Description: Isolation and molecular analysis of small GTP-binding proteins of Arabido
A;Reference number: JQ0907
A;Accession: JQ0910
A;Molecule type: mRNA
A;Residues: 2-196 <AN2>

C;Superfamily: ras transforming protein; translation elongation factor Tu homology
C;Keywords: GTP binding; membrane trafficking; nucleotide binding; P-loop
F;2-11//Domain: translation elongation factor Tu homology <ETU>
F;114-117/Region: GTP-binding NKXD motif
F;144-146/Region: nucleotide-binding motif A (P-loop)
F;14,15,33,114,115,117,144/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta

Query Match   13.3%;   Score 163.5;   DB 2;   Length 196;
Best Local Similarity 26.0%;   Pred. No. 1.6e-06;
Matches   50; Conservative   37; Mismatches   70; Indels   35; Gaps   6;

Qy   8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEE--KTYYIELW 65
            |:|:|||||||||:   : : :   :|| | |:| ::|:
Db   3 KLLLIGDSGVGKSCLLRFSDDSYVBSYISTIGV------DFKIRTVEQDGKTIKLQIW 55

Qy   66 DVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLV 125
            ||||:|| |||:||:::|||||||||| ::
Db   56 DTAGQ---ERFRTITSSYYRGAHGIIIVYDVTDESFNNVKQWLSEI------------ 99

Qy   126 TNGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNP--EEINLDCTNP 183
            :|: | |:   :|| |||:||| | ::||||||:|| |||:
Db   100 --------DRYASDNVNKLLVGNKSD-LTENRAIPYETAKAFADEIGIPFMETSAKDATNV 151

Qy   184 RYLAAGSSNAVK 195
            ||:|
Db   152 EQAFMAMSASIK 163

RESULT 9
B86153
ARA-5 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C;Accession: B86153
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: B86153
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-258 <STO>
C;Genetics:
A;Map position: 1
A;Cross-references: UNIPROT:P28188; GB:AE005172; NID:g2317906; PIDN:AAC24370.1; GSPDB:GN
C;Superfamily: ras transforming protein; translation elongation factor Tu homology

Query Match   13.3%;   Score 163.5;   DB 2;   Length 258;
Best Local Similarity 26.0%;   Pred. No. 2.4e-06;
Matches   50; Conservative   37; Mismatches   70; Indels   35; Gaps   6;

Qy   8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEE--KTYYIELW 65
            |:|:|||||||||:   : : :   :|| | |:| ::|:
Db   65 KLLLIGDSGVGKSCLLRFSDDSYVBSYISTIGV------DFKIRTVEQDGKTIKLQIW 117

Qy   66 DVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLV 125
            ||||:|| |||:||:::|||||||||| ::
Db   118 DTAGQ---ERFRTITSSYYRGAHGIIIVYDVTDESFNNVKQWLSEI------------ 161

Qy   126 TNGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNP--EEINLDCTNP 183
            :|: | |:   :|| |||:||| | ::||||||:|| |||:
Db   162 --------DRYASDNVNKLLVGNKSD-LTENRAIPYETAKAFADEIGIPFMETSAKDATNV 213

Qy   184 RYLAAGSSNAVK 195
            ||:|

Db   214 EQAFMAMSASIK 225

RESULT 10
A34716
GTP-binding protein SAS1 - slime mold (Dictyostelium discoideum)
C;Species: Dictyostelium discoideum
C;Date: 29-Jun-1990 #sequence_revision 29-Jun-1990 #text_change 09-Jul-2004
C;Accession: A34716; A61571
R;Saxe, S.A.; Kimmel, A.R.
Mol. Cell. Biol. 10, 2367-2378, 1990
A;Title: SAS1 and SAS2, GTP-binding protein genes in Dictyostelium discoideum with sequen
A;Reference number: A34716; MUID:90220623; PMID:2109188
A;Accession: A34716
A;Molecule type: mRNA
A;Residues: 1-208 <SAX>
A;Cross-references: UNIPROT:P20790; GB:M34456
R;Saxe, S.A.; Kimmel, A.R.
Dev. Genet. 9, 259-265, 1988
A;Title: Genes encoding novel GTP-binding proteins in Dictyostelium.
A;Reference number: A61571; MUID:89209367; PMID:3149563
A;Accession: A61571
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-208 <SA2>
C;Genetics:
A;Gene: SAS1
C;Superfamily: ras transforming protein; translation elongation factor Tu homology
C;Keywords: GTP binding; lipoprotein; membrane trafficking; nucleotide binding; P-loop; g
F;16-131/Domain: translation elongation factor Tu homology <ETU>
F;22-29/Region: nucleotide-binding motif A (P-loop)
F;128-131/Region: GTP-binding NKXD motif
F;158-160/Region: GTP-binding SAK/L motif
F;207,208/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match   13.2%;   Score 162.5;   DB 2;   Length 208;
Best Local Similarity 25.0%;   Pred. No. 2.2e-06;
Matches   51; Conservative   41; Mismatches   69; Indels   43; Gaps   8;

Qy   7 VKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTYYIELWD 66
            :|:||:||||||||| |:   ::
Db   16 IKLLLIGDSGVGKSCL--LLRPSEDSFTPSPITTIGIDFKIRTIE--LEGKRIKLQIWD 70

Qy   67 VGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLVT 126
            |||| || |||:||:::|||||||||| ::
Db   71 TAGQ---ERFRTITAYYRGAMGILLVDVTDEKSFGNIRNWI-----RNI---------- 113

Qy   127 NGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNPEEINLDCTNPRYL 186
            :|: | |:   :|| |||:||| | ::|||||||| :||:
Db   114 ------EQHATDSVNKMLIGNKCDMAE--KVVDSSRGKSLADEY--------GIKFL 155

Qy   187 AAGSSNAVKLSRFF---DKVIEKR 207
            | : |: ||||| |:|:
Db   156 ETSAKNSINVEAFISLAKDIKKR 179

RESULT 11
H71444
GTP-binding protein - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
A;Variety: columbia
C;Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 09-Jul-2004
C;Accession: H71444
R;Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirks
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.; Giele
avanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A;Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech,
erhoff, H.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ansc
C;Chalwatzis, N.
A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thali
A;Reference number: A71400; MUID:98121113; PMID:9461215
A;Accession: H71444

A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-221 <BEV>
A;Cross-references: UNIPROT:O23594; GB:Z97343; NID:g2245073; PIDN:CAB10533.1; PID:g2245451
C;Genetics:
A;Map position: 4COP9-4G3845
C;Superfamily: ras transforming protein; translation elongation factor Tu homology
C;Keywords: GTP binding; membrane trafficking; nucleotide binding; P-loop
P;28-143/Domain: translation elongation factor Tu homology <ETU>
P;34-41/Region: nucleotide-binding motif A (P-loop)
F;140-143/Region: GTP-binding NKXD motif
F;170-172/Region: GTP-binding SAK/L motif
F;40,41,59,140,141,143,170/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta

Query Match　13.2%;　Score 162;　DB 2;　Length 221;
Best Local Similarity　25.6%;　Pred. No. 2.6e-06;
Matches　50;　Conservative　37;　Mismatches　64;　Indels　44;　Gaps　8;

```
Qy    8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEB--KTYYIBLW  65
          |:::|||||||||  ||   ::           ::        |
Db   29 KLLLIGDSGVGKSCLLLRFADDSYLDSYISTIGV------DFKIRTVEQGKTIKLQIW   81

Qy   66 DVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLV 125
          || ||||  |||||  :||   :::|                    |
Db   82 DTAGQ---ERFRTITSSYYRGAHGIIVTYDVTDLESFNNVKQW----LN------- 123

Qy  126 TNGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNPBEINLDCTNPRY 185
          :::|| | |||   :: :    : :: :        |
Db  124 ----EIDRYASENVNKLLVGNKCDL--TSQKVVSTETAKAFAD----ELGIP------F 166

Qy  186 LAAGSSNAVKLSRFF 200
          ::       ||
Db  167 LETSAKNATNVEEAF 181
```

RESULT 12
S38083
probable purine nucleotide-binding protein YPT52 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YKR014c
C;Species: Saccharomyces cerevisiae
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
C;Accession: S38083
R;Duesterhoeft, A.; Moestl, D.; Poehlmann, R.; Philippsen, P.
submitted to the Protein Sequence Database, March 1994
A;Reference number: S37811
A;Accession: S38083
A;Molecule type: DNA
A;Residues: 1-234 <DUE>
A;Cross-references: UNIPROT:P36018; EMBL:Z28239; NID:g486428; PIDN:CAA82086.1; PID:g4864
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:YPT52
A;Cross-references: SGD:S0001722; MIPS:YKR014c
A;Map position: 11R
C;Superfamily: ras transforming protein; translation elongation factor Tu homology
C;Keywords: GTP binding; P-loop; purine nucleotide binding
F;10-17/Region: nucleotide-binding motif A (P-loop)
F;125-128/Region: GTP-binding NKXD motif
F;175-177/Region: GTP-binding SAK/L motif
F;16/Binding site: ATP/GTP (Lys) #status predicted

Query Match　13.2%;　Score 162;　DB 2;　Length 234;
Best Local Similarity　23.4%;　Pred. No. 2.8e-06;
Matches　55;　Conservative　37;　Mismatches　91;　Indels　52;　Gaps　7;

```
Qy    8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGC---SVDVRVHDYKEGTPEEKTYYIEL  64
          |:::||||||||| ::  ::           : :           |
Db    5 KLVLLGDSSVGKSSIVHRFVKDTDPELRESTIGAAFLSQSITIHPNDGNETKDVVIKFEI  64

Qy   65 WDVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVL 124
          |||||||||||||||||||||||||||||||            |
Db   65 WDTAGQ---ERYKSLAPMYYRNANAALVVYDITQEDSLQKARNWVDELKNK------- 112
```

Qy 125 VTNGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNPBEINLDCTNP- 183
```
          :::|:          ::      :
Db  113 ---------VGDDDLVIYLLGNKVDLCQETPS----TETSPDSNEGDEEQKVRAISTB 158

Qy  184 ------------RYLAAGSSNAVKLSRFFDKVIEKRYFL------REGNQIPG 218
              :          :        |
Db  159 EAKQYAQEQGLLFREVSAKTGEGVK--BIFQDIGEKLYDLKDEILSKQNRQIGG 211
```

RESULT 13
B34716
GTP-binding protein SAS2 - slime mold (Dictyostelium discoideum)
C;Species: Dictyostelium discoideum
C;Date: 29-Jun-1990 #sequence_revision 29-Jun-1990 #text_change 09-Jul-2004
R;Saxe, S.A.; Kimmel, A.R.
Mol. Cell. Biol. 10, 2367-2378, 1990
A;Title: SAS1 and SAS2, GTP-binding protein genes in Dictyostelium discoideum with seque
A;Reference number: A34716; MUID:90220623; PMID:2109188
A;Accession: B34716
A;Molecule type: DNA
A;Residues: 1-203 <SAX>
A;Cross-references: UNIPROT:P20791; GB:M34457
R;Saxe, S.A.; Kimmel, A.R.
Dev. Genet. 9, 259-265, 1988
A;Title: Genes encoding novel GTP-binding proteins in Dictyostelium.
A;Reference number: A61571; MUID:89209367; PMID:3149563
A;Accession: B61571
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-203 <SA2>
C;Genetics:
A;Gene: SAS2
C;Superfamily: ras transforming protein; translation elongation factor Tu homology
C;Keywords: GTP binding; lipoprotein; membrane trafficking; nucleotide binding; P-loop; l
F;16-131/Domain: translation elongation factor Tu homology <ETU>
F;22-29/Region: nucleotide-binding motif A (P-loop)
F;128-131/Region: GTP-binding NKXD motif
F;158-160/Region: GTP-binding SAK/L motif
F;202,203/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match　13.1%;　Score 161;　DB 2;　Length 203;
Best Local Similarity　25.2%;　Pred. No. 2.8e-06;
Matches　55;　Conservative　39;　Mismatches　76;　Indels　48;　Gaps　9;

```
Qy    7 VKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTYYIELWD  66
          |:|||||||||| ::  :             :        |
Db   16 VKLLLIGDSGVGKSCL--LLRPSDGSFTPSFIATIGIDFKIRTIE--LEGKRIKLQIWD  70

Qy   67 VGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLVT 126
          : |  :|  ::|                  |
Db   71 TAGQ---ERFRTITTAYYRGAMGILLVYDVTDEKSFGSIRNWI------RNI------- 113

Qy  127 NGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNPBEINLDCTNPRYL 186
          ::|| |::
Db  114 -----EQHASDSVNKMLIGNKCDMTB--KKVVDSSRGKSLADEY-----------GIKFL 155

Qy  187 AAGSSNAVKLSRFF--DKVIEKRYFLREGNQIPGFPD 221
          :|:  |   ::||
Db  156 ETSAKNSVNVEEAFIGLAKDIKKRMI----DTPNDPD 188
```

RESULT 14
T50323
ypt1-related protein 1 [imported] - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jul-2004
C;Accession: T50323
R;McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Cadieu, E.; Lelaure, V.; Galibert, F
submitted to the EMBL Data Library, January 2000
A;Reference number: Z25061
A;Accession: T50323
A;Status: preliminary; translated from GB/EMBL/DDBJ

```
A;Molecule type: DNA
A;Residues: 1-203 <MCD>
A;Cross-references: UNIPROT:P11620; EMBL:AL136535; PIDN:CAB66454.1; GSPDB:GN000067; SPDB:
A;Experimental source: strain 972h(-); cosmid c1703
C;Genetics:
A;Gene: SPDB:SPBC1703.10
A;Map position: 2
A;Introns: 4/2; 45/3; 61/3; 94/1
C;Superfamily: ras transforming protein; translation elongation factor Tu homology
F;9-124/Domain: translation elongation factor Tu homology <ETU>

Query Match     13.0%;  Score 160.5;  DB 2;  Length 203;
Best Local Similarity   23.7%;  Pred. No. 3.1e-06;
Matches    50;  Conservative   45;  Mismatches    67;  Indels   49;  Gaps    9;

Qy       8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEKTYYIELWDV 67
             |:|::|:||||||||   |     :   :: ::     ::     :||   :::||
Db      10 KLLLIGDSGVGKSCLLLRFADDTYTESYISTIGVDFKIRTFEL----EGKTVKLQIWDT 67

Qy      68 GGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRW------SLEALNR---- 116
             |:   :: : |||  :::   ||||||||||:||||::|||||        :: :||
Db      65 --ERFRTITSSYYRGAHGIIIVYDVTDQDSFNNVKQWLQEIDRYAVEGVNRLLVG 120

Qy     117 ---DLVPTGVLVTNGDYD-QEQFADN-QIPLLVI----GTKLDQIHETKRHEVLTRTAFL 167
              |:|  | :| ||:|:|     :|||  :|||||        ::|||||||:||:||||
Db     121 NKSDMVDKKVV----EYSVAKEFADSLNIPFLETSAKDSTNVEQAFLTMSRQIKERMG-- 174

Qy     168 AEDFNPEEINLDCTNPRYLAAGSSNAVKLSR 198
             : :: :::::|||:::::
Db     175 ------------NNTFASSNAKSSVKVGQ 191
```

```
RESULT 15
S04590
GTP-binding protein ypt1 - fission yeast (Schizosaccharomyces pombe)
N;Alternate names: ras-related protein ypt1
C;Species: Schizosaccharomyces pombe
C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 09-Jul-2004
C;Accession: S04590, S10025
R;Fawell, E.; Hook, S.; Armstrong, J.
Nucleic Acids Res. 17, 4373, 1989
A;Title: Nucleotide sequence of a gene encoding a YPT1-related protein from Schizosaccha
A;Reference number: S04590; MUID:89296485; PMID:2740223
A;Accession: S04590
A;Molecule type: DNA
A;Residues: 1-206 <FAW>
A;Cross-references: UNIPROT:P11620; EMBL:X15082; NID:g5145; PIDN:CAA33192.1; PID:g5146
R;Miyake, S.; Yamamoto, M.
EMBO J. 9, 1417-1422, 1990
A;Title: Identification of ras-related, YPT family genes in Schizosaccharomyces pombe.
A;Reference number: S10025; MUID:90228338; PMID:2328721
A;Accession: S10025
A;Molecule type: DNA
A;Residues: 'MNPE',8-206 <MIY>
A;Cross-references: EMBL:X52099; NID:g5147; PIDN:CAA36319.1; PID:g5148
C;Genetics:
A;Gene: ypt1
A;Introns: 48/3; 64/3; 97/1
C;Superfamily: ras transforming protein; translation elongation factor Tu homology
C;Keywords: GTP binding; lipoprotein; membrane trafficking; nucleotide binding; P-loop;
F;12-127/Domain: translation elongation factor Tu homology <ETU>
F;18-25/Region: nucleotide-binding motif A (P-loop)
F;124-127/Region: GTP-binding NKXD motif
F;154-156/Region: GTP-binding SAK/L motif
F;24,25,43,124,125,127,154/Binding site: Mg-GTP (Lys, Ser, Thr, Asn, Lys, Asp, Ser) #sta
F;205,206/Binding site: geranyl-geranyl (Cys) (covalent) #status predicted

Query Match     13.0%;  Score 160.5;  DB 2;  Length 206;
Best Local Similarity   23.7%;  Pred. No. 3.1e-06;
Matches    50;  Conservative   45;  Mismatches    67;  Indels   49;  Gaps    9;

Qy       8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEKTYYIELWDV 67
             |:|::|:||||||||   |     :   :: ::     ::     :||   :::||
```

Search completed: July 1, 2005, 00:00:51
Job time : 25 secs

```
              GenCore version 5.1.6
         Copyright (c) 1993 - 2005  Compugen Ltd.

OM protein - protein search, using sw model

Run on:    June 30, 2005, 19:57:06 ; Search time 84 Seconds
                                     (without alignments)
                                     1438.699 Million cell updates/sec

Title:     US-09-945-173-2
Perfect score: 1231
Sequence:  1 MASLDRVKVLVLGDSGVGKS..........PGFPDRKRFGAGTLKSLHYD 236

Scoring table:  BLOSUM62
           Gapop 10.0 , Gapext 0.5

Searched:  1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters:       1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    UniProt_03:*
           1: uniprot_sprot:*
           2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                        SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1216	98.8	236	2	Q8WUD3	Q8wud3 homo sapien
2	1133	92.0	236	2	Q9D4V7	Q9d4v7 m mus muscu
3	1132	92.0	236	2	Q8BMU2	Q8bmu2 m mus muscu
4	1085.5	88.2	235	2	Q6GPS4	Q6gps4 xenopus lae
5	1026.5	83.4	233	2	Q6TNS7	Q6tns7 brachydanio
6	641	52.1	129	2	Q9D0M6	Q9d0m6 m mus muscu
7	593.5	48.2	236	2	Q7Q5M9	Q7q5m9 anopheles g
8	502	40.8	274	2	Q9VXA9	Q9vxa9 drosophila
9	276	22.4	354	2	Q948D2	Q948d2 oryza sativ
10	276	22.4	354	2	Q7X985	Q7x985 oryza sativ
11	265	21.5	342	1	Y813_ARATH	Y813 arabidopsis
12	259	21.0	416	2	Q9U2V7	Q9u2v7 caenorhabdi
13	224	18.2	431	2	Q9FIB5	Q9fib5 arabidopsis
14	220.5	17.9	242	2	Q9VQG6	Q9vqg6 drosophila
15	189.5	15.4	208	2	Q75CA9	Q75ca9 ashbya goss
16	178.5	14.5	197	2	O18337	O18337 drosophila
17	175.5	14.3	206	2	Q9NFG0	Q9nfg0 plasmodium
18	173.5	14.1	197	2	Q9W4A0	Q9w4a0 drosophila
19	173	14.1	205	2	Q8TGD9	Q8tgd9 emericella
20	171.5	13.9	217	1	YPT2_VOLCA	P36861 volvox cart
21	171	13.9	203	2	Q41338	Q41338 lycopersico
22	170.5	13.9	401	2	Q9GNY1	Q9gny1 leishmania
23	170	13.8	203	2	O24112	O24112 nicotiana p
24	169.5	13.8	201	2	Q9W5X0	Q9w5x0 drosophila
25	169	13.7	202	2	Q40203	Q40203 lotus japon
26	169	13.7	203	1	RYL1_YARLI	P41924 yarrowia li
27	168.5	13.7	201	2	Q6P7Z7	Q6p7z7 xenopus tro
28	168.5	13.7	227	2	Q8AVL2	Q8avl2 xenopus lae
29	167.5	13.7	231	2	Q6NTJ2	Q6ntj2 xenopus lae
30	167.5	13.6	219	2	Q7Z160	Q7z160 toxoplasma
31	167	13.6	203	1	YPT2_MAIZE	Q05737 zea mays (m
32	167	13.6	203	2	Q41340	Q41340 lycopersico
33	167	13.6	203	2	Q9M7P5	Q9m7p5 capsicum an
34	167	13.6	209	2	Q9USG8	Q9usg8 tetrahymena
35	166.5	13.5	202	2	Q08155	Q08155 pisum sativ
36	166.5	13.5	207	2	Q7DLK9	Q7dlk9 vicia faba
37	166.5	13.5	207	2	Q6AW64	Q6aw64 entamoeba h
38	166	13.5	212	2	Q7QWP4	Q7wp4 giardia lam
39	166	13.5	212	2	Q9GU79	Q9gu79 giardia lam
40	165	13.4	202	2	Q9SXT5	Q9sxt5 cicer ariet
41	164.5	13.4	201	1	RB35_HUMAN	Q15286 homo sapien
42	164.5	13.4	201	2	Q6PHN9	Q6phn9 mus musculu
43	164.5	13.4	206	1	YPT5_CHLRE	Q39573 chlamydomon
44	164	13.3	192	2	Q9NDI3	Q9ndi3 entamoeba h
45	164	13.3	202	2	Q08153	Q08153 pisum sativ

```
                        ALIGNMENTS

RESULT 1
Q8WUD3
ID   Q8WUD3          PRELIMINARY;      PRT;    236 AA.
AC   Q8WUD3;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   RAB, member of RAS oncogene family-like 3.
GN   Name=RABL3;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.B., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Strausberg R.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC020832; AAH20832.1; -.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005525; F:GTP binding; IEA.
DR   GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.
DR   GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR   Pfam; PF00071; Ras; 1.
DR   PRINTS; PR00449; RASTRNSFRMNG.
DR   PROSITE; PS00675; SIGMA54_INTERACT_1; 1.
KW   GTP-binding.
SQ   SEQUENCE   236 AA;  26308 MW;  13D9D19F8FD96710 CRC64;

     Query Match             98.8%;  Score 1216;  DB 2;  Length 236;
```

Best Local Similarity 99.2%; Pred. No. 2.7e-88;
Matches 234; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy    1 MASLDRVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTY 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MASLDRVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTC 60

Qy   61 YIELWDVGGSVGSASSVKSTRAVFNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 YIELWDVGGSVGSASSVKSTRAVFNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVP 120

Qy  121 TGVLVTNGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNPEEINLDC 180
          ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db  121 TGVLVTNGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTTTAFLAEDFNPEEINLDC 180

Qy  181 TNPRYLAAGSSNAVKLSRFFDKVIEKRYFLREGNQIPGFPDRKRFPGAGTLKSLHYD 236
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 TNPRYLAAGSSNAVKLSRFFDKVIEKRYFLREGNQIPGFPDRKRFPGAGTLKSLHYD 236
```

RESULT 2
ID   Q9D4V7
AC   Q9D4V7;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Mus musculus adult male testis cDNA, RIKEN full-length enriched
DE   library, clone:493055305 product:hypothetical RAS small GTPases, Rab
DE   subfamily/ATP/GTP-binding site motif A (P-loop)/Sigma-54 factor
DE   interaction domain containing protein, full insert sequence (Rabl3
DE   protein).
GN   Name=Rabl3;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=21085560; PubMed=11217851; DOI=10.1038/35055500;
RA   RIKEN FANTOM Consortium;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,

RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RA   Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA   Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA   Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka Y., Hori F.,
RA   Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA   Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA   Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA   Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA   Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA   Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA   Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA   Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B5/EGFP transgenic ICR mice; TISSUE=Trophoblast Stem Cell;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [8]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B5/EGFP transgenic ICR mice; TISSUE=Trophoblast Stem Cell;
RA   Strausberg R.;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK016099; BAB30113.1; -.
DR   EMBL; BC050194; AAH50194.1; -.
DR   MGD; MGI:1914907; Rabl3.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005525; F:GTP binding; IEA.
DR   GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.
DR   GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR   InterPro; IPR001806; Ras trnsfrmng.
DR   InterPro; IPR002078; Sig54_interact.
DR   Pfam; PF00071; Ras; 1.
DR   PRINTS; PR00449; RASTRNSFRMNG.
DR   PROSITE; PS00675; SIGMA54_INTERACT_1; 1.
KW   GTP-binding; Hypothetical protein; Rabl3.
SQ   SEQUENCE 236 AA; 26297 MW; DFCB72737C11944B CRC64;

Query Match 92.0%; Score 1133; DB 2; Length 236;
Best Local Similarity 92.4%; Pred. No. 1e-81;
Matches 218; Conservative 3; Mismatches 15; Indels 0; Gaps 0;

```
Qy    1 MASLDRVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTY 60
```

```
Db          ||||| |||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
      1 MASLDRVKVLVLGDSGVGKSSLVHLLCHNQVLGNPSWTVGCSVDRVHDYKEGTPEEKTY 60

Qy     61 YIELWDVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVP 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 YIELWDVGGSVGSASSVKSTRAVFYNSVNGIIVHDLTNKKSSQNLYRWSLEVLNRDAVP 120

Qy    121 TGVLVTNGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNPEEINLDC 180
          |||||||||||       |||||||||||||||||||||||||||||||||||||||||
Db    121 TGVLVTNGDYDREQFADNQIPLLVIGTKLDQIHETKRHEVLIRTAFLAEDFNAEEINLDC 180

Qy    181 TNPRYLAAGSSNAVKLSRFFDKVIEKRYFLREGNQIPGFPDRKRFGAGTLKSLHYD 236
          ||||  |||||||||||||||||||| || |||||   |||||   :|:||
Db    181 TNPRSSAAGSSNAVKLSRFFDKVIEKRYFFREGNQIPGFSDRKRFGGGALKNFHCD 236

RESULT 3
Q8BMU2
ID   Q8BMU2_        PRELIMINARY;        PRT;    236 AA.
AC   Q8BMU2;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
DE   enriched library, clone:2610314P18 product:hypothetical RAS small
DE   GTPases, Rab subfamily/ATP/GTP-binding site motif A (P-loop)/Sigma-54
DE   factor interaction domain containing protein, full insert
DE   sequence.
GN   Name=Rab13;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Whole body;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Whole body;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   RIKEN FANTOM Consortium;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Whole body;
RA   The FANTOM Consortium;
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Whole body;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Whole body;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami K., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
```

```
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Whole body;
RA   Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA   Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA   Hayashida K., Hayatsu N., Hiramoto K., Hiraoka Y., Hirozane T.,
RA   Hori F., Inotani K., Ishii K., Ishii M., Kagawa I., Kasukawa T.,
RA   Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA   Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA   Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA   Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA   Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA   Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA   Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK028195; BAC25804.1; -.
DR   MGD; MGI:1914907; Rab13.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005525; F:GTP binding; IEA.
DR   GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.
DR   GO; GO:0000160; P:two-component signal transduction system (p.. .; IEA.
DR   InterPro; IPR001806; Ras_trnsfrmng.
DR   InterPro; IPR002078; Sig54_interact.
DR   Pfam; PF00071; Ras; 1.
DR   PRINTS; PR00449; RASTRNSFRMNG.
DR   PROSITE; PS00675; SIGMA54_INTERACT_1; 1.
KW   GTP-binding; Hypothetical protein.
SQ   SEQUENCE   236 AA;   26307 MW;   DFCE7271377C20DB CRC64;

Query Match             92.0%;  Score 1132;  DB 2;  Length 236;
Best Local Similarity   92.4%;  Pred. No. 1.2e-81;
Matches   218; Conservative   3; Mismatches  15; Indels   0; Gaps   0;

Qy     1 MASLDRVKVLVLGDSGVGKSSLVHLLCHNQVLGNPSWTVGCSVDRVHDYKEGTPEEKTY 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MASLDRVKVLVLGDSGVGKSSLVHLLCHNQVLGNPSWTVGCSVDRVHDYKEGTPEEKTY 60

Qy    61 YIELWDVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVP 120
         ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db    61 YIELWDVGGSVGSASSVKSTRAVFYNSVNGIILVHDLTNKKSSQNLYRWSLEVLNRDAVP 120

Qy   121 TGVLVTNGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNPEEINLDC 180
         |||||||||||      |||||||||||||||||||||||||||||||||||||||||
Db   121 TGVLVTNGDYDREQFADNQIPLLVIGTKLDQIHETKRHEVLIRTAFLAEDFNAEEINLDC 180

Qy   181 TNPRYLAAGSSNAVKLSRFFDKVIEKRYFLREGNQIPGFPDRKRFGAGTLKSLHYD 236
         ||||  ||||||||||||||||||||| || |||||  |||||   :|:|||
Db   181 TNPRSSAAGSSNAVKLSRFFDKVIEKRYFFREGNQIPGFSDRKRFGGGALKNFHCD 236

RESULT 4
Q6GPS4
ID   Q6GPS4_        PRELIMINARY;        PRT;    235 AA.
AC   Q6GPS4;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   MGC82648 protein.
GN   Name=MGC82648;
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
```

RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh L.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Rodrigues S., Sanchez A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.B., Schnerch A., Schein J.B.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA   Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA   Richardson P.;
RT   "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT   initiative.";
RL   Dev. Dyn. 225:384-391 (2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RA   Klein S., Gerhard D.S.;
RL   Submitted (JUN-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the small GTPase superfamily. Rab family.
DR   EMBL; BC073035; AAH73035.1; -.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005525; F:GTP binding; IEA.
DR   GO; GO:0015031; P:protein transport; IEA.
DR   GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR   InterPro; IPR003579; GTPase_Rab.
DR   InterPro; IPR001806; Ras_transfrmng.
DR   InterPro; IPR002078; Sig54_interact.
DR   Pfam; PF00071; Ras; 1.
DR   PRINTS; PR00449; RASTRNSFRMNG.
DR   SMART; SM00175; RAB; 1.
DR   PROSITE; PS00675; SIGMA54_INTERACT_1; 1.
KW   GTP-binding;
SQ   SEQUENCE 235 AA; 26180 MW; 9380A3E039C026CD CRC64;

Query Match   88.2%; Score 1085.5; DB 2; Length 235;
Best Local Similarity   90.3%; Pred. No. 5.8e-78;
Matches 213; Conservative   6; Mismatches 16; Indels   1; Gaps   1;

Qy   1 MASLDRVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTY 60
            ||||||||||||||||||||||||||||||||||||||||||||| ||: |:::|||||
Db   1 MASLDRVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRLHEYREGTPEEKTY 60

Qy   61 YIELWDVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVP 120
            |||||||||||||:|||||||||||||||:|||||||||||||||:||||||||||||
Db   61 YTELWDVGGSVGSASSVKSTRAVFYNAVNGIILVHDLTNKKSSQNLYRWSLEALNRDLQP 120

Qy   121 TGVLVTNGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNPEINLDC 180
            ||:||||||||::|:||||||||||||||||||:|:|||:||||||||||||:|:||||
Db   121 MGVLVTNGDYDRBQFADNQIPLLVIGTKLDQIPEAKRNEVLTRTAFLAEDFNABEINLDC 180

Qy   181 TNPRYLAAGSSNAVKLSRFFDKVIEKRYFLREGNOIPGPPDRKRFGGAGTLKSLHYD 236
            |||| ||||:|||||||:|||||||||:|| || ||:|||:||||||| ||:|||||
Db   181 TNTRCLAAGSSNAVKLSRFFPDKVIEKRY-PRBGNLIPGFSDRKRFGGGNFKSLHYD 235

RESULT 5

ID   Q6TNS7   PRELIMINARY;   PRT;   233 AA.
AC   Q6TNS7;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Hypothetical protein.
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Singapore local strain; TISSUE=Embryo;
RX   PubMed=12277932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.B., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Singapore local strain; TISSUE=Embryo;
RA   Strausberg R.;
RL   Submitted (JUL-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Belongs to the small GTPase superfamily. Rab family.
DR   EMBL; AY391440; AAQ91252.1; -.
DR   EMBL; BC078191; AAH78191.1; -.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005525; F:GTP binding; IEA.
DR   GO; GO:0015031; P:protein transport; IEA.
DR   GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.
DR   GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR   InterPro; IPR003579; GTPase_Rab.
DR   InterPro; IPR001806; Ras_transfrmng.
DR   InterPro; IPR002078; Sig54_interact.
DR   Pfam; PF00071; Ras; 1.
DR   PRINTS; PR00449; RASTRNSFRMNG.
DR   SMART; SM00175; RAB; 1.
DR   PROSITE; PS00675; SIGMA54_INTERACT_1; 1.
KW   GTP-binding; Hypothetical protein.
SQ   SEQUENCE 233 AA; 26113 MW; F14901939C884C086 CRC64;

Query Match   83.4%; Score 1026.5; DB 2; Length 233;
Best Local Similarity   83.5%; Pred. No. 2.7e-73;
Matches 197; Conservative 18; Mismatches 18; Indels   3; Gaps   1;

Qy   1 MASLDRVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTY 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db   1 MASLDRVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYREGTPEEKAF 60

RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Whole body;
RA   Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA   Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA   Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA   Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA   Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA   Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA   Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA   Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA   Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA   Tejima Y., Toya T., Yamamura Y., Yasunishi A., Yoshida K., Yoshino M.,
RA   Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK011280; BAB27514.1; -.
DR   MGD; MGI:1914907; Rabl3.
DR   GO; GO:0005525; F:GTP binding; IEA.
DR   GO; GO:0005525; F:DNA binding; IEA.
DR   GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.
DR   GO; GO:0000160; P:two-component signal transduction system (p. ..; IEA.
DR   InterPro; IPR001806; Ras_trnsfrmng.
DR   InterPro; IPR002078; Sig54_interact.
DR   Pfam; PF00071; Ras; 1.
DR   PRINTS; PR00449; RASTRNSFRMNG.
DR   PROSITE; PS00675; SIGMA54_INTERACT_1; 1.
KW   GTP-binding; Hypothetical protein.
SQ   SEQUENCE   129 AA; 14034 MW; 1E6B7723DF95499C CRC64;

Query Match   52.1%; Score 641; DB 2; Length 129;
Best Local Similarity   95.3%; Pred. No. 4.3e-43;
Matches 123; Conservative   1; Mismatches   5; Indels   0; Gaps   0;

Qy   1 MASLDRVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTY 60
           ||||||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db   1 MASLDRVKVLVLGDSGVGKSSLVHLLCHNQVLGNPSWTVGCSVDIRVHDYKEGTPEEKTY 60

Qy   61 YIELWDVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVP 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 YIELWDVGGSVGSASSVKSTRAVFYNSVNGIIIVHDLTNKKSSQNLYRWSLEVLNRDAVP 120

Qy   121 TGVLVTNGD 129
           |||||||||
Db   121 TGVLVTNGD 129

RESULT 7
Q7Q5M9
ID   Q7Q5M9   PRELIMINARY;   PRT;   236 AA.
AC   Q7Q5M9;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   AgCP5724 (Fragment).
GN   Name=agCG50989; ORFNames=ENSANGG00000010553;
OS   Anopheles gambiae str. PEST.
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX   NCBI_TaxID=180454;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=PEST;
RA   Anopheles Genome Sequencing Consortium;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
CC   EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC   preliminary data.
DR   EMBL; AAAB01008960; EAA10808.1; -.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005525; F:GTP binding; IEA.
DR   GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.

Qy   61 YIELWDVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVP 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||::|||
Db   61 YIELWDVGGSVGSASSVKSTRAVFYNSVNGIILVHDLTNKKSSQNLYRWSLEALSKDSSP 120

Qy   121 TGVLVTNGDYDQEQPADNQIPLLVIGTKLDQIHETKRHEVLTRTAPLABDFNPEEINLDC 180
           |||:::|||||||||||||||||||||||||||:::||:|||||||||:||:||||||||
Db   121 TGIIVSNGDYDREQPAENAVPLLLIGTKFDQIPENKRNDVLTRTAFLSBDFNABEINLDC 180

Qy   181 TNPRYLAAGSSNAVKLSRFPDKVIBKRYFLRBGNQIPGFPDRKRFGAGTLKSLHYD 236
           ||||||||||||||||||||||||||||||| : |: : |||:|:|| ||||
Db   181 TNPRYLAAGSSNAVKLSRFFDKVIBKRYFTRDPSQMQSFTDRRRF---NFKSLHSD 233

RESULT 6
Q9DOM6
ID   Q9DOM6   PRELIMINARY;   PRT;   129 AA.
AC   Q9DOM6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Mus musculus 10 days embryo whole body cDNA, RIKEN full-length
DE   enriched library, clone:2610002D21 product:hypothetical RAS small
DE   GTPases, Rab subfamily/ATP/GTP-binding site motif A (P-loop)/Sigma-54
DE   factor interaction domain containing protein, full insert
DE   sequence.
GN   Name=Rabl3;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Whole body;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Meth. Enzymol. 303:19-44(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Whole body;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   RIKEN FANTOM Consortium;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Whole body;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Whole body;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Konno H., Okazaki Y., Shibata K., Itoh M.,
RA   Carninci P., Shibata Y., Hayatsu M., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Whole body;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki P., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto K., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format

DR   GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR   InterPro; IPR000345; CytC_heme_BS.
DR   InterPro; IPR001806; Ras_trnsfrmng.
DR   InterPro; IPR002078; Sig54_interact.
DR   Pfam; PF00071; Ras; 1.
DR   PRINTS; PR00449; RASTRNSFRMNG.
DR   PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
DR   PROSITE; PS00675; SIGMA54_INTERACT_1; 1.
KW   GTP-binding.
FT   NON_TER   1
FT   NON_TER   236   236
SQ   SEQUENCE   236 AA;   26079 MW;   6FP2BA2D1D772A48 CRC64;

Query Match   48.2%;   Score 593.5;   DB 2;   Length 236;
Best Local Similarity   53.9%;   Pred. No. 5.3e-39;
Matches 118;   Conservative 41;   Mismatches 45;   Indels 15;   Gaps 5;

Qy   1 MASLDRVKVLVLGDSGVGKSSLVHLLCQNPWSWTVGCSVDVRVHDYKEGTPEEKTY 60
           |||||||||||||||||:||||| |:| ||||||| ||:|||:| |:|| |: ||:
Db   15 MAAIDKVRVLJVGDSGVGKTSLTHLIANNEPLTSPGWTVGCSVEVKLHEYREGTPAQNTF 74

Qy   61 YIELWDVGGSGVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVP 120
           ::||||||||| |:|||||| || |||||| ||||||||||||| |:||:| ||||:|
Db   75 FVELWDVGGSSI----SHKNTRGVFYNPTHGIILVHDLTNRKSQENLQRWLAEILNKD--- 127

Qy   121 TGVLVTNG----DYDQEQF-ADNQIPLLVIGTKLDQIHETKRHEVLTRTA--FLAEDFNP 173
           ||||||| |||||||| |||||||||||||||| ||||:|||: ||:| |||
Db   128 -GKDALKGGDAIDMDPEQFLGSTQIPILVIGTKLDMLDEGHKQKIQTRTSAGSIAEQCGA 186

Qy   174 EEINLDCTNPRYLAAGSSNAVKLSRFFDKVIEKRYFLRE 212
           |||||||||||||||||||:|:| ||||||||||:|:|:|
Db   187 DEICLNCHESRSLAAGTTDAVKLARFFDKVIERKYHNRD 225

RESULT 8
Q9VXA9
ID   Q9VXA9   PRELIMINARY;   PRT;   274 AA.
AC   Q9VXA9; Q8SZD5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   CG4789-PA (RE04047p).
GN   ORFNames=CG4789;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Liao X.,

RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodgeT, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426065; PubMed=12537568;
RA   Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA   Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA   George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA   Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA   Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA   Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT   "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT   melanogaster euchromatic genome sequence.";
RL   Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426070; PubMed=12537573;
RA   Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA   Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA   Ashburner M., Celniker S.E.;
RT   "The transposable elements of the Drosophila melanogaster euchromatin:
RT   a genomics perspective.";
RL   Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfied E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN   [5]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003503; AAF48668.2; -.
DR   EMBL; AY070945; AAL48567.1; -.
DR   FlyBase; FBgn0030792; CG4789.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005525; F:GTP binding; IEA.

DR   GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.
DR   GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR   InterPro; IPR002078; Sig54_interact.
DR   Pfam; PF00071; Ras; 1.
DR   PROSITE; PS00675; SIGMA54_INTERACT_1; 1.
KW   GTP-binding.
SQ   SEQUENCE   274 AA;   30049 MW;   6A0E65C28875A7CC CRC64;

Query Match   40.8%; Score 502; DB 2; Length 274;
Best Local Similarity   43.2%; Pred. No. 1.1e-31;
Matches 114; Conservative 45; Mismatches 67; Indels 38; Gaps 8;

Qy   1 MASLDRVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEKTY 60
Db   1 MAMNNRVRIVVVGDSGVGKTSLTHLITHNEALIRPGWTVGCNIQVKMHPFREGTARECPY 60

Qy   61 YIELWDVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRD--- 117
Db   61 FVELFDVGGSLNH----KNTRSVFYAGIDGIILVHDILTNAKSQRQLIDWLYEIVNKEGKD 116

Qy   118 ------LVP----------TGVLVTNGD-YDDEQF-ADNQIPLLVIGTKLDQIHETK 156
Db   117 TNKSNGASMPPSPPSPLSSFSTDNLGTDGHILPDMEEFLGATQTPILVMGTKLDLLDE-K 175

Qy   157 RHEVL--TRTAFLAEDFNPEEINLDCTNPRYLAAGSSNAVKLSRFPDKVIEKRYFLR--- 211
Db   176 RHPKMGVKKPGGIADKCCGAEEIWLNCRNSRSLAAGTTDAVKLSRFFDRVIENRKALRAAL 235

Qy   212 ---EGNQIPGFPDRKRFGAGTLK 231
Db   236 AFGVSSSNAVSPPDRRRFGFTSAK 259

RESULT 9
Q948D2
ID   Q948D2;   PRELIMINARY;   PRT;   354 AA.
AC   Q948D2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Putative GTP-BINDING PROTEIN YPTV3.
GN   Name=OSJNBa0013J21.21;
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wing R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA   Saski C., Henry D., Oates R., Simmons J.;
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC079843; AAL01181.1; -.
DR   Gramene; Q948D2; -.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005525; F:GTP binding; IEA.
DR   GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.
DR   GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR   InterPro; IPR001806; Ras_trnsfrmng.
DR   InterPro; IPR002078; Sig54_interact.
DR   Pfam; PF00071; Ras; 1.
DR   PRINTS; PR00449; RASTRNSFRMNG.
DR   PROSITE; PS00675; SIGMA54_INTERACT_1; 1.
KW   GTP-binding.
SQ   SEQUENCE   354 AA;   38719 MW;   DA15175E6A84A4A2 CRC64;

Query Match   22.4%; Score 276; DB 2; Length 354;
Best Local Similarity   29.8%; Pred. No. 1.2e-13;
Matches 75; Conservative 41; Mismatches 78; Indels 58; Gaps 9;

Qy   6 RVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVR-------VHDYKEGTP- 55

Db   27 QVRVLVVGDSGVGKSSLVHLILKGSAIARPPQTIGCAVDVKYIFIEVFPQFHITYGSPG 86

Qy   56 ------EEKTYIELWDVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSS 103

Db   87 SSSNSINSIKGDAERNFFVELWDVSGH---ERYKECRSLFYSQINGVFVYDLSQRKTK 142

Qy   104 QNLRRWSLEALNRDL--VPTGVLVTNGYDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVL 161

Db   143 TNLNKWAVEVAESGTFSAPLGSGGPGG--------LPVPYLVIANKVDIAPRDGKRVSS 193

Qy   162 TRTAFLAEDF--------NPEEINLDCTNPRYLAAGSSNAVKLSR------FPDKVIE 205

Db   194 GNLVDVARQVWBKQGLLPSSEELPLABSFPG--NSGLLTAAKVARYDKEALVKFFRMLIR 251

Qy   206 KRYFLREGNQIP 217

Db   252 RRYF--SNELP 260

RESULT 10
Q7X985
ID   Q7X985;   PRELIMINARY;   PRT;   354 AA.
AC   Q7X985;
DT   01-OCT-2003 (TrEMBLrel. 25, Created)
DT   01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Putative GTP-BINDING PROTEIN YPTV3.
GN   ORFNames=OSJNBa0013J21.21;
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   The Rice Chromosome 10 Sequencing Consortium;
RT   "In-depth view of structure, activity, and evolution of rice
RT   chromosome 10.";
RL   Science 300:1566-1569(2003).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
EMBL; AB017054; AAP52031.1; -.
DR   Gramene; Q7X985; -.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005525; F:GTP binding; IEA.
DR   GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.
DR   GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR   InterPro; IPR001806; Ras_trnsfrmng.
DR   InterPro; IPR002078; Sig54_interact.
DR   Pfam; PF00071; Ras; 1.
DR   PROSITE; PS00675; SIGMA54_INTERACT_1; 1.
KW   GTP-binding.
SQ   SEQUENCE   354 AA;   38719 MW;   DA15175E6A84A4A2 CRC64;

Query Match   22.4%; Score 276; DB 2; Length 354;
Best Local Similarity   29.8%; Pred. No. 1.2e-13;
Matches 75; Conservative 41; Mismatches 78; Indels 58; Gaps 9;

Qy   6 RVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVR-------VHDYKEGTP- 55

Db   6 RVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVR-------VHDYKEGTP- 55

Qy   27 QVRVLVVGDSGVGKSSLVHLILKGSAIARPPQTIGCAVDVKYIFIEVFPQFHITYGSPG 86

Qy   56 ------EEKTYIELWDVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSS 103

Db   87 SSSNSINSIKGDAERNFFVELWDVSGH---ERYKECRSLFYSQINGVFVYDLSQRKTK 142

Qy   104 QNLRRWSLEALNRDL--VPTGVLVTNGYDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVL 161

Db   143 TNLNKWAVEVAESGTFSAPLGSGGPGG--------LPVPYLVIANKVDIAPRDGKRVSS 193

Qy   162 TRTAFLAEDF--------NPEEINLDCTNPRYLAAGSSNAVKLSR------FFDKVIE 205

```
Db    194  GNLVDVARQWVEKQGLLPSSEELPLAESPG--NSGLLTAAKVARYDKEALVKFFRMLIR  251
           :|   |   :||.:|  |||:|||||||   |   -   :|  |||:|   || .:|

Qy    206  KRYFLREGNQIP  217
           :|||   |:::|

Db    252  RRYF--SNELP  260


RESULT 11
Y813_ARATH
ID   Y813_ARATH   STANDARD;      PRT;    342 AA.
AC   Q9C5J9; Q9LV96;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Hypothetical GTP-binding protein At5g64813.
GN   OrderedLocusNames=At5g64813; ORFNames=MXK3.3;
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20181125; PubMed=10718197;
RA   Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT   features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT   clones.";
RL   DNA Res. 7:31-63 (2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=22954850; PubMed=14593172; DOI=10.1126/science.1088305;
RA   Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA   Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA   Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA   Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA   Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA   Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA   Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA   Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
RA   Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA   Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA   Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
RA   Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT   "Empirical analysis of transcriptional activity in the Arabidopsis
RT   genome.";
RL   Science 302:842-846(2003).
CC   -!- CAUTION: Ref.1 sequence differs from that shown due to erroneous
CC       gene model prediction. The predicted gene should be splitted into
CC       3 genes: At5g64810, At5g64813 and At5g64816.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AB019236; BAA97293.1; ALT_SEQ.
DR   EMBL; AF360203; AAK25913.1; -.
DR   EMBL; AY040050; AAK64108.1; -.
DR   InterPro; IPR001806; Ras_trnsfrmng.
DR   Pfam; PF00071; Ras; 1.
DR   PRINTS; PR00449; RASTRNSFRMNG.
KW   GTP-binding; Hypothetical protein.
FT   DOMAIN        71      75       Poly-Ser.
FT   NP_BIND       29      36       GTP (Potential).
FT   NP_BIND       90      94       GTP (Potential).
FT   NP_BIND      160     163       GTP (Potential).
SQ   SEQUENCE   342 AA;  37741 MW;  3F26357637GAAF2E CRC64;

Query Match             21.5%;  Score 265;  DB 1;  Length 342;
Best Local Similarity   29.9%;  Pred. No. 8.6e-13;
Matches   73;  Conservative   42;  Mismatches   67;  Indels   62;  Gaps   10;

Qy      6  RVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTP----------   55
           :|||:||||:||||:||||||   :   .:   |||   .||||:||   :|.|   ::||
Db     22  QVRVLVVGDSGVGKTSLVHLINKGSSIVRPPQTIGCTVGVKHITY-GSPASSSSIQGD    79

Qy     56  EEKTYIELMVDVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALN  115
           ::   :   |   :||||||||   |||| :| |:|||||   |:   :   |::|   ||
Db     80  SERDFFVELMDVSGH----ERYKDCRSLFYSQINGVIFVHDLSQRRTKTSLQKWASE---  132

Qy    116  RDLVPTGV----LVTNGDYDQEQFADNQIPLLVIGTKLD----------QIHETKRHE   159
           |||   ||         |   ||.:   |||:   |:|   :||   |   ||||:||
Db    133  --VAATGTFSAPLPSCGP--------GGLPVPYIVVGNKADIAAKEGTKGSSGNLVDAARHW  184

Qy    160  VLTRTAF----------LAEDFNPEBINLDCTNPRYLAAGSS---NAVKLSRFFDKVIEKRY  208
           |||||              |   |||::||:   |||||   ||   |:   :   ||:|::
Db    185  VEKQGLLPSSSEDLPLFESFP--------GNGGLIAAAKETRYDKEALNKFFRMLIRRRY   236

Qy    209  FLRE  212
           |   |
Db    237  FSDE  240


RESULT 12
Q9U2V7
ID   Q9U2V7       PRELIMINARY;   PRT;    416 AA.
AC   Q9U2V7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein Y116A8C.10.
GN   ORFNames=Y116A8C.10;
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C. elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   McMurray A.A.;
RL   Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL117204; CAB55120.1; -.
DR   PIR; T31486; T31486.
DR   WormBase; WBGene00013790; Y116A8C.10.
DR   WormPep; Y116A8C.10; CE23321.
DR   GO; GO:0005525; F:GTP binding; IEA.
DR   GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.
DR   InterPro; IPR001623; DnaJ_N.
DR   InterPro; IPR001806; Ras_trnsfrmng.
DR   PRINTS; PR00449; RASTRNSFRMNG.
KW   Hypothetical protein.
SQ   SEQUENCE   416 AA;  47639 MW;  630A1B79022E8A34 CRC64;

Query Match             21.0%;  Score 259;  DB 2;  Length 416;
Best Local Similarity   32.2%;  Pred. No. 3.3e-12;
Matches   76;  Conservative   32;  Mismatches   86;  Indels   42;  Gaps   10;

Qy      1  MASLDR--VKVLVLGDSGVGKSSLVHLLCQNQVLGNP---------SWTVGCSVDVRVHDYK   51
           |   |||   |:|||||||:|||||   :||||||   :||   :|   |||:||||:||:
Db    162  MSSLDESSTKILVLGDSCVGKTSLCHCIAGG---GEPPGGGRSFDSTIGATVMAWHEYR  218
```

```
Qy      52  EGTPERKTYYIELWDVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRW--  109
                ||||||:|:|| ||:|||:|||||:|:|       ||:   ||||||||||:||
Db     219  AGTPEQRTELLELWDIGGMVAHRQAAQ----VFFEGAVGAILVHDLTNKRSEENLATWLT  274

Qy     110  SLEALNRDLVPTGVLVTNGDYDQEQFADNQIPLLVIGTKLDQI-HE----TKRHEVLTRT  164
                |:  |    ||::|||||:|||: |:| ||: |||||| :      ::::
Db     275  MLDGKPRGAAPKSSKDPAAVALKVDIESCNIPVLIGTKADLIVPHKGPVSYDRYDGLGNA  334

Qy     165  AFLAEDFNPEBINLDCTNPRYLAAGSSNAVKLSRFPDKVIEK------RYFLREG       213
                |||                      :|:|||           :|:|
Db     335  ---------NSAISVWKPRNRKRKNSKKPMKIGR--KLVVEKEKVKKFGRQYLRRG       378
```

RESULT 13

Q9FIB5
```
ID   Q9FIB5          PRELIMINARY;     PRT;    431 AA.
AC   Q9FIB5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MYH9.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9156233; PubMed=10048488;
RA   Asamizu E., Sato S., Kaneko T., Nakamura Y., Kotani H., Miyajima N.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. VIII.
RT   Sequence features of the regions of 1,081,958 bp covered by seventeen
RT   physically assigned P1 and TAC clones.";
RL   DNA Res. 5:379-391(1998).
DR   EMBL; AB016893; BAB09412.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR003822; PAH.
SQ   SEQUENCE  431 AA;  48540 MW;  FFBD3C8C7D1D1167 CRC64;
```
```
Query Match             18.2%;  Score 224;  DB 2;  Length 431;
Best Local Similarity   27.5%;  Pred. No. 2e-09;
Matches   66;  Conservative   40;  Mismatches   66;  Indels   68;  Gaps   11;

Qy      6  RVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGT------PEB   57
               ::|||:||| ||:| | |    |   :  ::|| ||:|   ||::|
Db     22  QIRVLVVGDSG---SSIV--------RPSQTIGCTVGVKHLTYASPASSSSIIKGDSE   68

Qy     58  KTYYIELWDVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLRALNRD  117
               : ::|||||||       :|::|| ||||:  :|||||||||||:|:|:
Db     69  RDFFVELWDVSGH----ERYKDCRSLFYSQINGVIFVHDLSQRTKTNLQKWA-------  117

Qy    118  LVPTGVLVTNGDYDQEQFADN---QIPLLVIGTKLD------------QIHETKRHEVL  161
               :|       ||||||:         :|:||||||            ||||:|||||
Db    118  ---GEVSVTGEFSAPLSSGGPGGLPVPYIVIGNKADIAAKGGTNGSSGNLVDAARHWVE  173

Qy    162  TRTAF---------LAEDFNPEBINLDCTNPRYLAAGSSNAVK--LSRFFDKVIEKRYFLRE  212
               ::|           |||||||:||||||||:||||||:|||    |:||||||:|||:||
Db    174  KQGLLPHSDELPLSBSF-PSNVGL-----IMAAKEARYDKEALTKIFHMLIRRRYFSDE  226
```

RESULT 14

Q9VQG6
```
ID   Q9VQG6          PRELIMINARY;     PRT;    242 AA.
AC   Q9VQG6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   CG15399-PA.
GN   ORFNames=CG15399;
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed-10731132; DOI=10.1126/science.287.5461.2185;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nixon K.A., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zhong X.H., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426065; PubMed=12537558;
RA   Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA   Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA   George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA   Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA   Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA   Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT   "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT   melanogaster euchromatic genome sequence.";
RL   Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426070; PubMed=12537573;
RA   Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA   Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA   Ashburner M., Celniker S.E.;
RT   "The transposable elements of the Drosophila melanogaster euchromatin:
RT   a genomics perspective.";
RL   Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfied B.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
```

```
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN   [5]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RG   FlyBase;
RL   Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003582; AAF51207.1; -.
DR   FlyBase; FBgn0031460; CG15399.
SQ   SEQUENCE   242 AA;   28362 MW;  3260907A8B357727 CRC64;

Query Match             17.9%;  Score 220.5;  DB 2;  Length 242;
Best Local Similarity   27.1%;  Pred. No. 1.9e-09;
Matches   64;  Conservative   41;  Mismatches   76;  Indels   55;  Gaps    8;

Qy         7 VKVLVLGDSGVGKSSLVHLLCQNQVLGNP-SWTVG---CSVDVRVHDYKE-----GTPEE 57
             |:::|||||  |||:|||   :.:   ||:|::   :            |
Db        12 VRILMLGDRGVGKTSLTNLMATTEITPTPDSRTVGEESWHVQVRLHEYSKPVILPPTPTW 71

Qy        58 KT------------------YYIELMDVGGSVGSASSVKSTRAVFYNSVNGIIFV 94
             :|             |:|:.   : ||:|:
Db        72 TTPSSEDSENYPYMRSTPTTNILYFVERYDLN--SDWRMCRQQRESFYKNIDGIVLV 128

Qy        95 HDLTNKKSSQNLRRWSLEAL-----NRDLVPTGVLVTNGYDYDQEQFADNQIPLLVIGTKL 149
             ..:    |:.|  |.  : :        .||
Db       129 YNMLELSSQDSLHDWLYDPLRQICKHRHLRIRSIL---------KNHNAPILVVGTNL 177

Qy       150 DQIHETKRHEVLTRTAFLAEDFNPEEINLDCTNPRYLAAGSSNAVKLSRFPDKVIE 205
             | |||  :| |  | | | :.:: :|||
Db       178 DKL-----MRRPLRRGSIAHQLNVBEMLVNCLDPQSFVDKSRNQGKLYGFLNRVIE 229

RESULT 15
ID   Q75CA9       PRELIMINARY;      PRT;   208 AA.
AC   Q75CA9;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   ACR003Cp.
GN   ORFNames=ACR003C;
OS   Ashbya gossypii (Yeast) (Eremothecium gossypii).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Eremothecium.
OX   NCBI_TaxID=33169;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 10895;
RA   Brachat S., Voegeli S.R., Dietrich F.S., Lerch A., Gaffney T.,
RA   Philippsen P.;
RL   Submitted (SEP-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE016816; AAS51230.1; -.
DR   HSSP; P10114; 1KAO.
DR   AGD; ACR003C; -.
DR   GO; GO:0003677; F:DNA binding; IEA.
DR   GO; GO:0005525; F:GTP binding; IEA.
DR   GO; GO:0006886; P:intracellular protein transport; IEA.
DR   GO; GO:0007264; P:small GTPase mediated signal transduction; IEA.
DR   GO; GO:0000160; P:two-component signal transduction system (p. .; IEA.
DR   InterPro; IPR003579; GTPase_Rab.
DR   InterPro; IPR003577; GTPase_Ras.
DR   InterPro; IPR003578; GTPase_Rho.
DR   InterPro; IPR002041; RAN.
DR   InterPro; IPR001806; Ras trnsfrmng.
DR   InterPro; IPR002078; Sig54 interact.
DR   InterPro; IPR005225; Small_GTP.
DR   Pfam; PF00071; Ras; 1.
DR   PRINTS; PR00449; RASTRNSFRMNG.
DR   SMART; SM00175; RAB; 1.
DR   SMART; SM00176; RAN; 1.
DR   SMART; SM00173; RAS; 1.
DR   SMART; SM00174; RHO; 1.
DR   TIGRFAMs; TIGR00231; small_GTP; 1.
DR   PROSITE; PS00675; SIGMA54_INTERACT_1; 1.
KW   GTP-binding.
SQ   SEQUENCE   208 AA;   22970 MW;  D83DC3A96C43E6CB CRC64;

Query Match             15.4%;  Score 189.5;  DB 2;  Length 208;
Best Local Similarity   27.5%;  Pred. No. 4.4e-07;
Matches   55;  Conservative   40;  Mismatches   66;  Indels   39;  Gaps    9;

Qy         7 VKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDY--KEGTPE-EKTYIE 63
             :|:::|||||||||:|||  ::  :|:  ::
Db         9 LKVIILGDSGVGKTSLMHRYVNDKYSQQYKATIGA------DFLTKEVTVDGDKVATMQ 61

Qy        64 LWDVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGV 123
             :| || |||||||||||| |::||||  :|:|:||  :|:::
Db        62 VWDTAGQ----ERFQSLGVAFYRGADCCVLVYDVTNAKSFENVKSWRDEFLVHANVPS-- 115

Qy       124 LVTNGYDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNPEEINLDCTNP 183
             ||||  |||   |
Db       116 --------PFSF------PFVILGNKID-VEESKK-VVTTRTS----------- 152

Qy       184 RYLAAGSSNAVKLSRFPDKV 203
             :  ||||
Db       153 PFFLTSAKNAINVDDAFEEI 172

Search completed: July  1, 2005, 00:00:26
Job time : 91 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 23:29:44 ; Search time 29 Seconds
(without alignments)
607.489 Million cell updates/sec

Title: US-09-945-173-2
Perfect score: 1231
Sequence: 1 MASLDRVKVLVLGDSGVGKS........PGFPDRKRFGAGTLKSLHYD 236

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :   Issued Patents AA:*
1: /cgn2_6/ptcodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptcodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptcodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptcodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptcodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptcodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	450	36.6	325	4	US-09-270-767-45478	Sequence 45478, A
2	166.5	13.5	227	4	US-09-270-767-46812	Sequence 46812, A
3	164.5	13.4	228	4	US-09-949-016-9506	Sequence 9506, Ap
4	163	13.2	201	2	US-08-531-525-13	Sequence 13, Appl
5	163	13.2	201	2	US-08-718-270A-13	Sequence 13, Appl
6	156.5	12.7	201	3	US-08-916-901-8	Sequence 8, Appli
7	156.5	12.7	201	3	US-09-154-602-8	Sequence 8, Appli
8	156	12.7	228	4	US-09-270-767-46283	Sequence 46283, A
9	155.5	12.6	203	4	US-09-255-920A-12	Sequence 12, Appl
10	154.5	12.6	201	3	US-08-916-901-3	Sequence 3, Appli
11	154.5	12.6	201	3	US-09-154-602-3	Sequence 3, Appli
12	154.5	12.6	202	2	US-08-531-525-14	Sequence 14, Appl
13	154.5	12.6	202	2	US-08-718-270A-14	Sequence 14, Appl
14	154.5	12.6	205	4	US-09-709-103-49	Sequence 49, Appl
15	154.5	12.6	205	4	US-09-439-410A-49	Sequence 49, Appl
16	154.5	12.6	237	4	US-09-949-016-7643	Sequence 7643, Ap
17	154	12.5	197	4	US-09-270-767-45823	Sequence 45823, A
18	153	12.4	206	4	US-09-828-310-13	Sequence 13, Appl
19	152.5	12.4	215	2	US-08-531-525-10	Sequence 10, Appl
20	152.5	12.4	215	2	US-08-718-270A-10	Sequence 10, Appl
21	150.5	12.2	238	4	US-09-248-796A-19962	Sequence 19962, A
22	149.5	12.1	202	2	US-08-531-525-24	Sequence 24, Appl
23	149.5	12.1	202	2	US-08-718-270A-24	Sequence 24, Appl
24	149	12.1	207	2	US-08-824-873-4	Sequence 4, Appli
25	149	12.1	207	3	US-09-198-184-4	Sequence 4, Appli
26	148	12.0	227	4	US-09-270-767-43285	Sequence 43285, A
27	148	12.0	227	4	US-09-248-796A-20291	Sequence 20291, A
28	147.5	12.0	342	4	US-09-949-016-10797	Sequence 10797, A
29	146.5	11.9	131	4	US-09-270-767-58626	Sequence 58626, A
30	146.5	11.9	214	4	US-09-270-767-33012	Sequence 33012, A
31	146.5	11.9	214	4	US-09-270-767-48229	Sequence 48229, A
32	145	11.8	369	4	US-09-949-016-7355	Sequence 7355, Ap
33	144.5	11.7	191	2	US-08-531-525-26	Sequence 26, Appl
34	144.5	11.7	191	2	US-08-718-270A-26	Sequence 26, Appl
35	143.5	11.7	203	4	US-09-949-016-9379	Sequence 9379, Ap
36	143.5	11.6	214	3	US-08-741-411-11	Sequence 11, Appl
37	142.5	11.6	205	2	US-08-531-525-25	Sequence 25, Appl
38	142.5	11.6	205	2	US-08-718-270A-25	Sequence 25, Appl
39	142.5	11.6	213	4	US-09-248-796A-20293	Sequence 20293, A
40	142	11.5	128	3	US-09-513-999C-5775	Sequence 5775, Ap
41	142	11.5	194	3	US-08-741-411-9	Sequence 9, Appli
42	142	11.5	203	4	US-09-949-016-6440	Sequence 6440, Ap
43	142	11.5	213	2	US-08-531-525-12	Sequence 12, Appl
44	142	11.5	213	2	US-08-718-270A-12	Sequence 12, Appl
45	142	11.5	216	4	US-09-949-016-7056	Sequence 7056, Ap

ALIGNMENTS

RESULT 1
US-09-270-767-45478
; Sequence 45478, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45478
; LENGTH: 325
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-45478

Query Match   36.6%; Score 450; DB 4; Length 325;
Best Local Similarity   42.1%; Pred. No. 7.1e-42;
Matches 104; Conservative 40; Mismatches 65; Indels 38; Gaps 8;

Qy   18 GKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEKTYYIELWDVGGSVGSASSV 77
             |||:|| ||: .:|| ||:::|||| ::|: ::|||:|||| |::|||||
Db   70 GKTSLTHLITHNEALIRPGWTVGCNIQVKMHPFREGTARECPYFVELFDVGGSLNH---- 125

Qy   78 KSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRD--------------LVP------ 120
             :||||| ::||||||||||| ||| ::|:
Db   126 KNTRSVFYAGIDGIILVHDLTNAKSQRQLIDWLYEIVNKEGKDTNKSNGASMPPSPPSPL 185

Qy   121 ----TGVLVTNGD--YDQEQF-ADNQIPLLVIGTKLDQIHETKRHEVL--TRTAPLAEDF 171
                 |||: ||:   :|:|:| ::||| :|| ||| ||: : || | ||:
Db   186 SSFSTDNLGTDGHILFDMEEFLGATQTPILVMGTKLDLLDE-KRHPKMGVKKPGGIADKC 244

Qy   172 NPEEINLDCTNPRYLAAGSSNAVKLSRFPDKVIEKRYFLR-------EGNQIPGFPDRKR 224
             |||| | :||||: ||||||||| :|||||| :|| |:||||
Db   245 GAEEIWLNCRNSRSLAAGTTDAVKLSRFFDRVIENRKALRAALAFGVSSSNAVSPPDRRR 304

Qy   225 FGAGTLK 231
             ||:|:
Db   305 FGPTSAK 311

RESULT 2
US-09-270-767-46812
; Sequence 46812, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:

```
Db          182 VNVEEM-FNCITELVLRAKKDNLAK 205
```

RESULT 4
US-08-531-525-13
; Sequence 13, Application US/08531525
; Patent No. 5840683
; GENERAL INFORMATION:
;   APPLICANT: Hlavka, Joseph J.
;   APPLICANT: Pincus, Matthew R.
;   APPLICANT: No. 5840683le, John F.
;   APPLICANT: Abajian, Henry B.
;   APPLICANT: Kende, Andrew S.
;   TITLE OF INVENTION: Peptides Inhibiting the Oncogenic Action
;   TITLE OF INVENTION: of P21 Ras
;   NUMBER OF SEQUENCES: 52
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Greenlee and Winner, P.C.
;   STREET: 5370 Manhattan Circle, Suite 201
;   CITY: Boulder
;   STATE: Colorado
;   COUNTRY: US
;   ZIP: 80303
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/531,525
;   FILING DATE: 21-SEP-1995
;   CLASSIFICATION: 530
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Ferber, Donna M.
;   REGISTRATION NUMBER: 33,878
;   REFERENCE/DOCKET NUMBER: 37-94
;   TELECOMUNICATION INFORMATION:
;   TELEPHONE: (303) 499-8080
;   TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 13:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 201 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   ORIGINAL SOURCE:
;   ORGANISM: Discopyge ommata
US-08-531-525-13

Query Match   13.2%; Score 163; DB 2; Length 201;
Best Local Similarity 25.9%; Pred. No. 6.1e-10;
Matches   51; Conservative 42; Mismatches 80; Indels 24; Gaps 6;

Qy   8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTYYIELWDV 67
          |:|:||||||||||: | : ||: |:| ::||
Db   10 KLLLIGDSGVGKSCLLLRFADDTYTSYISTIGVDFKIRTIEL----DGKTIKLQIWDT 64

Qy   68 GGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRW------SLRALNRDLVP 120
          || | : | || :: ::|||: ||||| :|
Db   65 AGQ---ERFRTITSSYYRGAHGIIVVYDVTDQESFNNVKQWLQEIDRYASENVNKLLVG 120

Qy 121 TGVLVTNG---DYDQEQFADN-QIPLLVI---GTKLDQIHETKRHEVLTRTAFLAEDFN 172
          : ||::   ::|: |:|
Db 121 NKCDLTTKKVVDYTTKEFADSLGIFLETSAKNATNVEQAFMTMAAEIKKRMGPGATSGG 180

Qy 173 PEEINLDCTNPRYLAAG 189
          ::|: ||:||
Db 181 SEKSNVNIQSTPVKSSG 197

RESULT 5
```

APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 46812
; LENGTH: 227
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-46812

Query Match   13.5%; Score 166.5; DB 4; Length 227;
Best Local Similarity 31.5%; Pred. No. 3e-10;
Matches   40; Conservative 25; Mismatches 53; Indels .9; Gaps 2;

Qy   8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTYYIELWDV 67
          |:|:||||||||||: | : ||: |:| ::||
Db   36 KLLLIGDTGVGKSSLLIRFSDDTFSGSYITTIGVDFKIRTVDI----EGMRVKLQIWDT 90

Qy   68 GGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLVTN 127
          || | : | || :: ::|||: |||| :|
Db   91 AGQ---ERFRTITSTYYRGTHGVIVVYDVTNGESFANVRRWLEBIQNNCDVVKKVLVGN 146

Qy 128 GDYDQEQ 134
          : ||:
Db 147 KNDDPDR 153

RESULT 3
US-09-949-016-9506
; Sequence 9506, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9506
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-9506

Query Match   13.4%; Score 164.5; DB 4; Length 228;
Best Local Similarity 25.9%; Pred. No. 5.1e-10;
Matches   53; Conservative 36; Mismatches 63; Indels 53; Gaps 9;

Qy   8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPB---EKTYYIEL 64
          |:|:||||||||||: | : ||: |:| ::||
Db   37 KLLLIGDSGVGKSSLLRFADNTFSGSYITTIGV------DFKIRTVEINGEKV-KLQI 88

Qy   65 WDVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVL 124
          || :: |: | || :: ::|||: ||||| :|
Db   89 WDTAGQ---ERFRTITSTYYRGTHGVIVVYDVTSAESFVNVKRW-LHEINQNC------ 137

Qy 125 VTNGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAF------------LAED 170
          : ||::   : ::|:
Db 138 -------------DDVCRIIVGNKND---DPERKVVETEDAYKFAGQMGIQLFETSAKEN 181

Qy 171 FNPEEINLDCTNPRYLAAGSSNAVK 195
          ||: ||: :|
```

US-08-718-270A-13
; Sequence 13, Application US/08718270A
; Patent No. 5910478
; GENERAL INFORMATION:
;   APPLICANT: Hlavka, Joseph J.
;   APPLICANT: Pincus, Matthew R.
;   APPLICANT: No. 5910478le, John F.
;   APPLICANT: Abajian, Henry B.
;   APPLICANT: Kende, Andrew S.
;   TITLE OF INVENTION: Peptidomimetics Inhibiting
;   TITLE OF INVENTION: the Oncogenic Action of P21 Ras
;   NUMBER OF SEQUENCES: 52
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
;     STREET: 5370 Manhattan Circle, Suite 201
;     CITY: Boulder
;     STATE: Colorado
;     COUNTRY: US
;     ZIP: 80303
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/718,270A
;     FILING DATE: 20-SEP-1996
;     CLASSIFICATION: 514
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 08/531,525
;     FILING DATE: 21-SEP-1995
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 60/004,091
;     FILING DATE: 21-SEP-1995
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Ferber, Donna M.
;     REGISTRATION NUMBER: 33,878
;     REFERENCE/DOCKET NUMBER: 78-95
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (303) 499-8080
;     TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 13:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 201 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   ORIGINAL SOURCE:
;     ORGANISM: Discopyge ommata
US-08-718-270A-13

Query Match      13.3%;  Score 163; DB 2; Length 201;
Best Local Similarity   25.9%;  Pred. No. 6.1e-10;
Matches   51;  Conservative  42;  Mismatches  80;  Indels  24;  Gaps   6;

Qy        8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTYYIELWDV 67
            |:||:||||||||    :    | |    :   :.:  ||  | :||:||
Db       10 KLLLIGDSGVGKSCLLLRFADDTYTESYISTIGVDFKIRTIEL-----DGKTIKLQIWDT 64

Qy       68 GGSVGSASSVKSTRAVFVNSVNGIIFVHDLTNKKSSQNLRRW------SLEALNRDLVP 120
            |    :  . :  :  :| |||: ||||:||||:       SLEALNRDLVP
Db       65 AGQ---ERFRTITSSYYRGAHGIIVVYDVTDQESFNNVKQWLQBIDRYASENVNKLLVG 120

Qy      121 TGVLVTNG---DYDQEQFADN-QIPLLVI----GTKLDQIHETKRHEVLTRTAFLABDFN 172
            :        |||:||    |:          :|:
Db      121 NKCDLLTKKVVDYTTKEFADSLGIPFLETSAKNATNVEQAFMTMAAEIKKRMGPGATSGG 180

Qy      173 PEEINLDCTNPRYLAAG 189
            | :|: |||   ::|
Db      181 SEKSNVNIQSTPVKSSG 197

RESULT 6
US-08-916-901-8
; Sequence 8, Application US/08916901
; Patent No. 5892012
; GENERAL INFORMATION:
;   APPLICANT: Hillman, Jennifer L.
;   APPLICANT: Lal, Preeti
;   APPLICANT: Corley, Neil C.
;   APPLICANT: Shah, Purvi
;   TITLE OF INVENTION: RAB PROTEINS
;   NUMBER OF SEQUENCES: 9
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Incyte Pharmaceuticals, Inc.
;     STREET: 3174 Porter Dr.
;     CITY: Palo Alto
;     STATE: CA
;     COUNTRY: USA
;     ZIP: 94304
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Diskette
;     COMPUTER: IBM Compatible
;     OPERATING SYSTEM: DOS
;     SOFTWARE: FastSEQ for Windows Version 2.0
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/916,901
;     FILING DATE: Filed Herewith
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Billings, Lucy J.
;     REGISTRATION NUMBER: 36,749
;     REFERENCE/DOCKET NUMBER: PF-0367 US
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 415-855-0555
;     TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 201 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   IMMEDIATE SOURCE:
;     LIBRARY: GenBank
;     CLONE: 57006
US-08-916-901-8

Query Match      12.7%;  Score 156.5;  DB 2;  Length 201;
Best Local Similarity   24.1%;  Pred. No. 3.3e-09;
Matches   49;  Conservative  42;  Mismatches  69;  Indels  43;  Gaps   8;

Qy        8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTYYIELWDV 67
            |:||:||||||||    :    | | :    : .:  ||  |:| DGKTIKLQIWDT 64
Db       10 KLLLIGDSGVGKSCLLLRFADDTYTESYISTIGVDFKIRTIEL-----DGKTIKLQIWDT 64

Qy       68 GGSVGSASSVKSTRAVFVNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLVTN 127
            |    :  . :  :  :| |||: ||||:||:::|| :|| :|              106
Db       65 AGQ---ERFRTVTSSYYRGAHGIIVVYDVTDQESYANVKQWLQEI----            106

Qy      128 GDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLABDFNPEBINLDCTNPRYLA 187
            :|| : |   :||||| | |    | | :::                           |:|
Db      107 ----DRYASENVNKLLVGNKSDL--TTKK--VVDNTT--AKEF------ADSLGVPFLE 149

Qy      188 AGSSNAVKLSRFPDKV---IEKR 207
            :   :  |  :|:|        ::|
Db      150 TSAKNATNVEQAFMTMAAEIKKKR 172

RESULT 7
US-09-154-602-8
; Sequence 8, Application US/09154602
; Patent No. 6300472

Fri Jul 1 07:44:54 2005

us-09-945-173-2.ral

```
;   GENERAL INFORMATION:
;     APPLICANT: Hillman, Jennifer L.
;     APPLICANT: Lal, Preeti
;     APPLICANT: Corley, Neil C.
;     APPLICANT: Shah, Purvi
;     TITLE OF INVENTION: RAB PROTEINS
;     NUMBER OF SEQUENCES: 9
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Incyte Pharmaceuticals, Inc.
;       STREET: 3174 Porter Dr.
;       CITY: Palo Alto
;       STATE: CA
;       COUNTRY: USA
;       ZIP: 94304
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE: Diskette
;       COMPUTER: IBM Compatible
;       OPERATING SYSTEM: DOS
;       SOFTWARE: FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/09/154,602
;       FILING DATE:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 08/916,901
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME: Billings, Lucy J.
;       REGISTRATION NUMBER: 36,749
;       REFERENCE/DOCKET NUMBER: PF-0367 US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 415-855-0555
;       TELEFAX: 415-845-4166
;   INFORMATION FOR SEQ ID NO: 8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 201 amino acids
;       TYPE: amino acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;     IMMEDIATE SOURCE:
;       LIBRARY: GenBank
;       CLONE: 57006
US-09-154-602-8

Query Match    12.7%;   Score 156.5; DB 3;   Length 201;
Best Local Similarity   24.1%;   Pred. No. 3.3e-09;
Matches  49;   Conservative  42;   Mismatches  69;   Indels  43;   Gaps   8;

Qy     8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTYYIELWDV  67
           |:||:|||||||:|: |:| ::|  | :|  |  |:  | ::|
Db    10 KLLLIGDSGVGKSCLLLRFADDTYTESYISTIGVDFKIRTIEL----DGKTIKLQIWDT  64

Qy    68 GGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLVTN 127
            :: :| ||:|:| |:||||||| :|:||:|:| |:::|
Db    65 AGQ---ERFRTVTSSYYRGAHGIIVVYDVTDQESYANVAQWLQEI--------------- 106

Qy   128 GDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNPEEINLDCTNPRYLA 187
           :|:  :|  |||||:|:|||||::: |||| |:
Db   107 ----DRYASENVNKLLVGNKSDL--TTKK--VVDNTT--AKEF-----ADSLGVPFLE 149

Qy   188 AGSSNAVKLSRFFDKV---IEKR 207
            :: ::|| |:|| |: |:||
Db   150 TSAKNATNVEQAFWTMAAEIKKR 172

RESULT 8
US-09-270-767-46283
;   Sequence 46283, Application US/09270767
;     Patent No. 6703491
;   GENERAL INFORMATION:
;     APPLICANT: Homburger et al.
;     TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;     FILE REFERENCE: File Reference: 7326-094
;     CURRENT APPLICATION NUMBER: US/09/270,767
```

```
;     CURRENT FILING DATE: 1999-03-17
;     NUMBER OF SEQ ID NOS: 62517
;     SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 46283
;     LENGTH: 228
;     TYPE: PRT
;     ORGANISM: Drosophila melanogaster
US-09-270-767-46283

Query Match    12.7%;   Score 156;  DB 4;   Length 228;
Best Local Similarity   27.5%;   Pred. No. 4.6e-09;
Matches  44;   Conservative  30;   Mismatches  56;   Indels  30;   Gaps   5;

Qy     8 KVLVLGDSGVGKSSLVHLLCQNQV--LGNPSWTVGCSVDVRVHDYKEGTPEEKTYYIELW  65
           ::::|||||||:: | |  :
Db    21 RLILIGDSTVGKSSLLKFFTDGKFAELSDP--TVGVDFFARLIEMKDGTQIK----IQLW  74

Qy    66 DVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLV 125
           | |:  :| || :|:  :|:|||||:: | ||:|:::| |   |:| :::|
Db    75 DTAGQERFRSITKS----YYRNSVGVLLVVDISNHASFEHIPLWMMEA----------- 118

Qy   126 TNGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTA 165
            ::  |||: ||:|| |::|||| |:|| || |
Db   119 ------QRHIEPHRPVPALVGCKLDLINAGGHREVTTEEA 152
```

```
RESULT 9
US-09-255-920A-12
;   Sequence 12, Application US/09255920A
;     Patent No. 6623980
;   GENERAL INFORMATION:
;     APPLICANT: Fisher, Joseph
;     APPLICANT: Lorens, James
;     APPLICANT: Anderson, David
;     APPLICANT: Luo, Ying
;     APPLICANT: Huang, Betty
;     APPLICANT: Shen, Mary
;     TITLE OF INVENTION: EXO1 and EXO2, EXOCYTOTIC PROTEINS
;     FILE REFERENCE: A65905-1/DJB/RMS
;     CURRENT APPLICATION NUMBER: US/09/255,920A
;     CURRENT FILING DATE: 1999-02-23
;     PRIOR APPLICATION NUMBER: 60/075,534
;     PRIOR FILING DATE: 1998-02-23
;     PRIOR APPLICATION NUMBER: 60/086,650
;     PRIOR FILING DATE: 1998-05-26
;     NUMBER OF SEQ ID NOS: 12
;     SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 12
;     LENGTH: 203
;     TYPE: PRT
;     ORGANISM: Mouse
;     FEATURE:
;       NAME/KEY: UNSURE
;       LOCATION: (34)
;       OTHER INFORMATION: The xaa at position 34 represents an unknown amino
;       OTHER INFORMATION: acid.
US-09-255-920A-12

Query Match    12.6%;   Score 155.5; DB 4;   Length 203;
Best Local Similarity   23.8%;   Pred. No. 4.3e-09;
Matches  51;   Conservative  39;   Mismatches  69;   Indels  55;   Gaps   9;

Qy     8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSW--TVGCSVDVRVHDYKEGTPEEKTYYIELW  65
           |:||:|||||||:: | |  :|   | ::|
Db    10 KLLLIGDSGVGKTCVLFRFSED-AFXNSTFISTIGIDFKIRTIEL----DGKRIKLQIW  63

Qy    66 DVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLV 125
           | |:  :| || :|:  :|:|||||:: | ||:|:::| |   |:|:  |
Db    64 DTAGQ----ERFRTITTAYYRGAMGIMLVYDITNEKSFDNIRXNWI------RNI---- 107

Qy   126 TNGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFN-------PEEIN 177
            ::  |||: ||:|| |::|||| |:|| || |
Db   108 ------EEHASADVERMILGNKCDV--NDKRQVSKERGEKALDYGIKFMETSAKANIN 158
```

Qy   178 LDCT---------NPRYLAAGSSNAVKLS 197
              :: |:: ||||:||
Db   159 VENAFFTLARDIKAKMDKNWKATAAGSSHGVKIT 192

RESULT 10
US-08-916-901-3
; Sequence 3, Application US/08916901
; Patent No. 5892012
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: RAB PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Incyte Pharmaceuticals, Inc.
;   STREET: 3174 Porter Dr.
;   CITY: Palo Alto
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94304
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/916,901
;   FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Billings, Lucy J.
;   REGISTRATION NUMBER: 36,749
;   REFERENCE/DOCKET NUMBER: PF-0367 US
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-855-0555
;   TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 201 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; IMMEDIATE SOURCE:
;   LIBRARY: LIVRTUT04
;   CLONE: 2514506
US-08-916-901-3

Query Match   12.6%; Score 154.5; DB 2; Length 201;
Best Local Similarity   24.1%; Pred. No. 5.5e-09;
Matches 49; Conservative 42; Mismatches 69; Indels 43; Gaps   8;

Qy   8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTYYIELWDV 67
              |:|::||||||||: | :||| :| |:||:|
Db   10 KLLLIGDSGVGKSCLLLRFADDTYTESYISTIGVDFKIRTIEL-----DGKTIKLQIWDT 64

Qy   68 GGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLVTN 127
              :: :|: : ||| |||::|:|:| |||:||
Db   65 AGQ---ERFRTITSSYYRGAHGIIVVDVTDQESYANVKQWLQEI------------- 106

Qy   128 GDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNPEEINLDCTNPRYLA 187
              :|: | |||:| ||:|:
Db   107 -----DRYASENVNKLLVGNKSDL--TTKK--VVDNNT--AKEF------ADSLGIPFLE 149

Qy   188 AGSSNAVKLSRFFDKV---IEKR 207
              || | | |:|
Db   150 TSAKNATNVEQAFMTMAAEIKKR 172

RESULT 11
US-09-154-602-3
; Sequence 3, Application US/09154602
; Patent No. 6300472
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: RAB PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Incyte Pharmaceuticals, Inc.
;   STREET: 3174 Porter Dr.
;   CITY: Palo Alto
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94304
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/154,602
;   FILING DATE:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/916,901
;   FILING DATE:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Billings, Lucy J.
;   REGISTRATION NUMBER: 36,749
;   REFERENCE/DOCKET NUMBER: PF-0367 US
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-855-0555
;   TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 201 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; IMMEDIATE SOURCE:
;   LIBRARY: LIVRTUT04
;   CLONE: 2514506
US-09-154-602-3

Query Match   12.6%; Score 154.5; DB 3; Length 201;
Best Local Similarity   24.1%; Pred. No. 5.5e-09;
Matches 49; Conservative 42; Mismatches 69; Indels 43; Gaps   8;

Qy   8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTYYIELWDV 67
              |:|::||||||||: | :||| :| |:||:|
Db   10 KLLLIGDSGVGKSCLLLRFADDTYTESYISTIGVDFKIRTIEL-----DGKTIKLQIWDT 64

Qy   68 GGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLVTN 127
              :: :|: : ||| |||::|:|:| |||:||
Db   65 AGQ---ERFRTITSSYYRGAHGIIVVDVTDQESYANVKQWLQEI------------- 106

Qy   128 GDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNPEEINLDCTNPRYLA 187
              :|: | |||:| ||:|:
Db   107 -----DRYASENVNKLLVGNKSDL--TTKK--VVDNNT--AKEF------ADSLGIPFLE 149

Qy   188 AGSSNAVKLSRFFDKV---IEKR 207
              || | | |:|
Db   150 TSAKNATNVEQAFMTMAAEIKKR 172

RESULT 12
US-08-531-525-14
; Sequence 14, Application US/08531525
; Patent No. 5840683

GENERAL INFORMATION:
APPLICANT: Hlavka, Joseph J.
APPLICANT: Pincus, Matthew R.
APPLICANT: No. 5840683le, John F.
APPLICANT: Abajian, Henry B.
APPLICANT: Kende, Andrew S.
TITLE OF INVENTION: Peptides Inhibiting the Oncogenic Action
TITLE OF INVENTION: of P21 Ras
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
STREET: 5370 Manhattan Circle, Suite 201
ADDRESSEE: Greenlee and Winner, P.C.
CITY: Boulder
STATE: Colorado
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/531,525
FILING DATE: 21-SEP-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Ferber, Donna M.
REGISTRATION NUMBER: 33,878
REFERENCE/DOCKET NUMBER: 37-94
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 202 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Lymnea stagnalis
US-08-531-525-14

Query Match   12.6%;   Score 154.5;   DB 2;   Length 202;
Best Local Similarity   22.8%;   Pred. No. 5.6e-09;
Matches 44; Conservative 38; Mismatches 70; Indels 41; Gaps 5;

```
Qy    8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTYYIELWDV 67
          |:|::||||||||| |:  :|  :::||
Db   13 KLLLIGDSGVGKSCLLRFADDTYTESYISTIGVDFKIRTIEL-----DGKTIKLQIWDT 67

Qy   68 GGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLVTN 127
          :|||:|:|::||
Db   68 AGQ---ERFRTITSSYYRGAHGIIVVYDVTDQESFNNVKQWLQEI----------- 109

Qy  128 GDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNPEEINLDCTNPRYLA 187
          :|:||   :: |:
Db  110 -----DRYASENVNKLLVGNKSDL--TTKK---------VDFTTAKEYADQLGIPFLE 151

Qy  188 AGSSNAVKLSRFF 200
          : :|:|
Db  152 TSAKNATNVEQAF 164
```

RESULT 13
US-08-718-270A
; Sequence 14, Application US/08718270A
; Patent No 5910478
; GENERAL INFORMATION:
; APPLICANT: Hlavka, Joseph J.
; APPLICANT: Pincus, Matthew R.
; APPLICANT: No. 591047Ble, John F.

; APPLICANT: Abajian, Henry B.
; APPLICANT: Kende, Andrew S.
; TITLE OF INVENTION: Peptidomimetics Inhibiting
; TITLE OF INVENTION: the Oncogenic Action of P21 Ras
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/718,270A
; FILING DATE: 20-SEP-1996
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/531,525
; FILING DATE: 21-SEP-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/004,091
; FILING DATE: 21-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 78-95
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 202 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Lymnea stagnalis
US-08-718-270A-14

Query Match   12.6%;   Score 154.5;   DB 2;   Length 202;
Best Local Similarity   22.8%;   Pred. No. 5.6e-09;
Matches 44; Conservative 38; Mismatches 70; Indels 41; Gaps 5;

```
Qy    8 KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTYYIELWDV 67
          |:|::||||||||| |:  :|  :::||
Db   13 KLLLIGDSGVGKSCLLRFADDTYTESYISTIGVDFKIRTIEL-----DGKTIKLQIWDT 67

Qy   68 GGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLVTN 127
          :|||:|:|::||
Db   68 AGQ---ERFRTITSSYYRGAHGIIVVYDVTDQESFNNVKQWLQEI----------- 109

Qy  128 GDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNPEEINLDCTNPRYLA 187
          :|:||   :: |:
Db  110 -----DRYASENVNKLLVGNKSDL--TTKK---------VDFTTAKEYADQLGIPFLE 151

Qy  188 AGSSNAVKLSRFF 200
          : :|:|
Db  152 TSAKNATNVEQAF 164
```

RESULT 14
US-09-709-103-49
; Sequence 49, Application US/09709103
; Patent No. 6733991
; GENERAL INFORMATION:
; APPLICANT: Cismowski, Mary

Search completed: July 1, 2005, 00:02:36
Job time : 31 secs

```
;  APPLICANT: Duzic, Emir
;  TITLE OF INVENTION: AGS Proteins and Nucleic Acid Molecules and Uses Therefor
;  FILE REFERENCE: 60388-A-PCT-US
;  CURRENT APPLICATION NUMBER: US/09/709,103
;  CURRENT FILING DATE: 2000-11-08
;  NUMBER OF SEQ ID NOS: 73
;  SOFTWARE: PatentIn version 3.1
;  SEQ ID NO 49
;   LENGTH: 205
;   TYPE: PRT
;   ORGANISM: Homo Sapiens
US-09-709-103-49

Query Match         12.6%;  Score 154.5;  DB 4;  Length 205;
Best Local Similarity   22.7%;  Pred. No. 5.7e-09;
Matches  46;  Conservative  41;  Mismatches  73;  Indels  43;  Gaps  6;

Qy    8  KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTYYIELWDV   67
            |:|::|||||||||  :      :     :|:|  :|    ||   ||  ::||
Db   13  KLLLIGDSGVGKSCLLLRFADDTYTESYISTIGVDFKIRTIEL-----DGKTIKLQIWDT   67

Qy   68  GGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLVTN  127
            ::  :| |  :|  :|||:  :    :||::|::|:|    |
Db   68  AGQ----ERFRTITSSYYRGAHGIIVYDVTDQESFNNVKQWLQEI------------   109

Qy  128  GDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNPEEINLDCTNPRYLA  187
            ::|  :| |:|  ::|:|     |:   :|
Db  110  ----DRYASENVNKLLVGNKCDL--TTKK---------VVDYTTAKEFADSLGIPFLE  152

Qy  188  AGSSNAVKLSRFFDKV---IEKR  207
            :  || : :| :  :||
Db  153  TSAKNATNVEQSFMTMAAEIKKR  175


RESULT 15
US-09-439-410A-49
;  Sequence 49, Application US/09439410A
;  Patent No. 6746852
;  GENERAL INFORMATION:
;  APPLICANT: Cismowski, Mary
;  APPLICANT: Duzic, Emir
;  TITLE OF INVENTION: AGS PROTEIN AND NUCLEIC ACID MOLECULES AND USES THEREOF
;  FILE REFERENCE: 1919/60388-B
;  CURRENT APPLICATION NUMBER: US/09/439,410A
;  CURRENT FILING DATE: 1999-11-11
;  NUMBER OF SEQ ID NOS: 118
;  SOFTWARE: PatentIn version 3.1
;  SEQ ID NO 49
;   LENGTH: 205
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-439-410A-49

Query Match         12.6%;  Score 154.5;  DB 4;  Length 205;
Best Local Similarity   22.7%;  Pred. No. 5.7e-09;
Matches  46;  Conservative  41;  Mismatches  73;  Indels  43;  Gaps  6;

Qy    8  KVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVDVRVHDYKEGTPEEKTYYIELWDV   67
            |:|::|||||||||  :      :     :|:|  :|    ||   ||  ::||
Db   13  KLLLIGDSGVGKSCLLLRFADDTYTESYISTIGVDFKIRTIEL-----DGKTIKLQIWDT   67

Qy   68  GGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKSSQNLRRWSLEALNRDLVPTGVLVTN  127
            ::  :| |  :|  :|||:  :    :||::|::|:|    |
Db   68  AGQ----ERFRTITSSYYRGAHGIIVYDVTDQESFNNVKQWLQEI------------   109

Qy  128  GDYDQEQFADNQIPLLVIGTKLDQIHETKRHEVLTRTAFLAEDFNPEEINLDCTNPRYLA  187
            ::|  :| |:|  ::|:|     |:   :|
Db  110  ----DRYASENVNKLLVGNKCDL--TTKK---------VVDYTTAKEFADSLGIPFLE  152

Qy  188  AGSSNAVKLSRFFDKV---IEKR  207
            :  || : :| :  :||
Db  153  TSAKNATNVEQSFMTMAAEIKKR  175
```

THIS PAGE BLANK (USPTO)

```
                    GenCore version 5.1.6
            Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        June 30, 2005, 13:28:23 ; Search time 2650.06 Seconds
                                         (without alignments)
                                         10212.469 Million cell updates/sec

Title:         US-09-945-173-3
Perfect score: 711
Sequence:      1 atggcgtccctggatcgggt.........agagccttcattatgactga 711

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      3439544 seqs, 1903213400 residues

Total number of hits satisfying chosen parameters:    68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     EST:*
               1: gb_est1:*
               2: gb_est2:*
               3: gb_htc:*
               4: gb_est3:*
               5: gb_est4:*
               6: gb_est5:*
               7: gb_est6:*
               8: gb_gss1:*
               9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	686.2	96.5	918	4	BI862084	BI862084 603390970
2	684	96.2	718	5	BX099226	BX099226 BX099226
3	682.2	95.9	720	4	BG501635	BG501635 602548678
4	681.4	95.8	732	6	BI333788	BI333788 602999279
5	680.6	95.7	699	7	CN361072	CN361072 170006000
6	679.4	95.6	720	7	CN361073	CN361073 328775509
7	674.8	94.9	815	5	BQ215951	BQ215951 AGENCOURT
8	668	94.0	1867	3	CR619977	CR619977 full-leng
9	663.6	93.3	797	1	AL516309	AL516309 AL516309
10	654.4	92.0	989	4	BM451532	BM451532 AGENCOURT
11	650.2	91.4	902	4	BG258348	BG258348 602379905
12	650	91.4	855	7	CK455329	CK455329 915906 MA
13	644	90.6	785	4	BI759699	BI759699 603045561
14	636.4	89.5	649	4	BG025782	BG025782 602274775
15	628	88.3	959	4	BG031602	BG031602 602299747
16	613.6	86.3	696	7	CN794203	CN794203 4129345 B
17	611	85.9	986	6	BI334986	BI334986 602998879
18	589.8	83.0	920	1	AA203636	AA203636 zx58b11.r
19	588.8	82.8	602	4	BG613092	BG613092 602640623
20	587.8	82.7	993	3	AK016099	AK016099 Mus muscu
21	586.2	82.4	2013	3	AK028195	AK028195 Mus muscu
22	583	82.0	584	7	CN361075	CN361075 170004241
c 23	581	81.7	875	7	CK456385	CK456385 917058 MA
24	576.8	81.1	2006	3	AK028269	AK028269 Mus muscu
25	573.8	80.7	573	7	CF730316	CF730316 UI-M-GZ0-
26	562.6	79.1	665	4	BG080166	BG080166 H3050E03-
27	559	78.6	786	2	BF106997	BF106997 601824352
28	558	78.5	827	2	BF571796	BF571796 602078477
29	548.8	77.2	722	2	BB610950	BB610950 BB610950
30	542.4	76.3	743	2	BB611463	BB611463 BB611463
31	540	75.9	551	1	AA662211	AA662211 ns66d03.r
32	534.4	75.2	810	6	CD512097	CD512097 AGENCOURT
33	533.4	75.0	536	7	CN361079	CN361079 170005977
34	531	74.7	573	2	AW373567	AW373567 PM1-BT053
35	525.6	73.9	1053	5	BM015053	BM015053 603641023
36	514	72.3	526	5	BU431648	BU431648 UI-HF-BN0
37	511	71.9	518	4	BM147898	BM147898 TCAAP1Q15
38	507	71.3	563	2	BF208017	BF208017 601862341
39	503.6	70.8	826	1	AJ456809	AJ456809 AJ456809
40	497.6	70.0	798	6	CD298863	CD298863 AGENCOURT
41	491.6	69.1	598	5	BU706219	BU706219 UI-M-FR0-
42	483.4	68.0	494	7	CN361078	CN361078 170005319
43	481.2	67.7	800	5	BU460217	BU460217 603368513
44	478.2	67.3	573	6	CA542843	CA542843 C0624F08-
45	477.4	67.1	536	4	BG610366	BG610366 326250 MA

ALIGNMENTS

```
RESULT 1
BI862084
LOCUS       BI862084                 918 bp    mRNA    linear    EST 10-OCT-2001
DEFINITION  603390970F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:5399969 5',
            mRNA sequence.
ACCESSION   BI862084
VERSION     BI862084.1  GI:16002831
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 918)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12019  row: 1  column: 18
            High quality sequence stop: 833.
FEATURES             Location/Qualifiers
     source          1..918
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5399969"
                     /tissue_type="mammary adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_87"
                     /note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.383 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
ORIGIN

Query Match             96.5%;    Score 686.2;  DB 4;  Length 918;
Best Local Similarity   99.3%;    Pred. No. 4e-189;
Matches 700; Conservative 0; Mismatches 3; Indels 2; Gaps 1;
```

```
                          bin/showLib.pl.cgi/response?libNo=972  Contact: Ina Rolfs
                          RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
                          Heubnerweg 6, D-14059 Berlin, Germany
                          Tel: +49 30 32639 101
                          Fax: +49 30 32639 111
                          www.rzpd.de
                          This clone is available royalty-free from RZPD;
                          contact RZPD (clone@rzpd.de) for further information. Seq primer:
                          M13r, Primer sequence: TTTCACACAGAAACAGCTATGAC.
FEATURES          Location/Qualifiers
     source           1..718
                          /organism="Homo sapiens"
                          /mol_type="mRNA"
                          /db_xref="taxon:9606"
                          /clone="IMAGp998I061023 ; IMAGE:435149"
                          /sex="male"
                          /dev_stage="20 week-post conception fetus"
                          /lab_host="DH10B (ampicillin resistant)"
                          /clone_lib="Soares fetal liver-spleen_1NFLS_S1"
                          /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                          with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                          This is a subtracted version of the original Soares fetal
                          liver spleen 1NFLS library.    1st strand cDNA was primed
                          with a Pac I - oligo(dT) primer [5'
                          AACTGGAAGAATTAATTAAGAGATCTTTTTTTTTTTTTTTTTTTT 3'],
                          double-stranded cDNA was ligated to Eco RI adaptors
                          (Pharmacia), digested with Pac I and cloned into the Pac I
                          and Eco RI sites of the modified pT7T3 vector. Library
                          went through one round of normalization. Library
                          constructed by Bento Soares and M.Fatima Bonaldo."
ORIGIN
    Query Match      96.2%;  Score 684;  DB 5;  Length 718;
    Best Local Similarity  100.0%;  Pred. No. 1.6e-188;
    Matches  684;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        28 CTGGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAAT 87
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 CTGGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAAT 60

Qy        88 CAAGTGCTCGGGAAATCCATCATGGACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGAT 147
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 CAAGTGCTCGGGAAATCCATCATGGACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGAT 120

Qy       148 TACAAAGAAGGAACCCAGAAGAAGAAGACCTACTACATAGAATTATTGGGATGGTTGGAGGC 207
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 TACAAAGAAGGAACCCAGAAGAAGAAGACCTACTACATAGAATTATTGGGATGGTTGGAGGC 180

Qy       208 TCTGTGGGCAGTGCCAGCAGCGTGAAAAGCACAAGAGCAGCAGTATTCTACAACTCCGTAAAT 267
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 TCTGTGGGCAGTGCCAGCAGCGTGAAAAGCACAAGAGCAGCAGTATTCTACAACTCCGTAAAT 240

Qy       268 GGTATTATTTTCGTACACGACTTAACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTGTGG 327
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 GGTATTATTTTCGTACACGACTTAACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTGTGG 300

Qy       328 TCATTGGAAGCTCTCAACAGGGATTTGGTGCCAACTGGACGCTTGGTGACAAATGGGGAT 387
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 TCATTGGAAGCTCTCAACAGGGATTTGGTGCCAACTGGACGCTTGGTGACAAATGGGGAT 360

Qy       388 TATGATCAAGAACAGTTGCTGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTG 447
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 TATGATCAAGAACAGTTGCTGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTG 420

Qy       448 GACCAGATTCATGAAACAAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAG 507
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 GACCAGATTCATGAAACAAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAG 480

Qy       508 GATTTCAATCCAGAAGAAATTAATTTGGACTGCACAAATCCACGGTACTTAGCTGCAGGT 567
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 GATTTCAATCCAGAAGAAATTAATTTGGACTGCACAAATCCACGGTACTTAGCTGCAGGT 540

Qy       568 TCTTCCAATGCTGTCAAGCTCAGTAGGTTTTTGATAAGGTCATAGAAGAAGAGATACTTT 627
```

(rotated left-margin sequence block:)

```
Qy         9 CCTGGATCGCGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTTAGT 68
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         7 CCTGGATCGCGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTTAGT 66

Qy        69 CCATCTCCTATGCCAAAATCAAGTGCTCGGGAAATCCATCATGGACTGTGGGCTGCTCAGT 128
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        67 CCATCTCCTATGCCAAAATCAAGTGCTCGGGAAATCCATCATGGACTGTGGGCTGCTCAGT 126

Qy       129 GGATGTCAGAGTTCATGATTACAAAGAAGGAACCCAGAAGAAGAAGACCTACTACATAGA 188
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       127 GGATGTCAGAGTTCATGATTACAAAGAAGGAACCCAGAAGAAGAAGACCTACTACATAGA 186

Qy       189 ATTATGGGATGGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGAAAAGCACAAGAGCAGT 248
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       187 ATTATGGGATGGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGAAAAGCACAAGAGCAGT 246

Qy       249 ATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTAACAAATAAGAAGTCCTC 308
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       247 ATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTAACAAATAAGAAGTCCTC 306

Qy       309 CCAAAACTTGCGTCGTGTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCAACTGGAGT 368
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       307 CCAAAACTTGCGTCGTGTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCAACTGGAGT 366

Qy       369 CTTGGTGACAAATGGGATTATGATCAAGAACAGTTGCTGATAACCAAATACCACTGTT 428
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       367 CTTGGTGACAAATGGGATTATGATCAAGAACAGTTGCTGATAACCAAATACCACTGTT 426

Qy       429 GGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTTTAACTAG 488
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       427 GGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTTTAACTAG 486

Qy       489 GACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGCACAAATCC 548
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       487 GACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGCACAAATCC 546

Qy       549 ACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCAGTAGGTTTTTGATAAGGT 608
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       547 ACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCAGTAGGTTTTTGATAAGGT 606

Qy       609 CATAGAGAAGAGATACTTTTTAAGAGAAGGTAA--TCAGATTCCAGGCTTCCTGATCGG 666
             |||||||||||||||||||||||||||||||      |||||||||||||||||||||||
Db       607 CATAGAGAAGAGATACTTTTTAAGAGAAGGTAAATCACGATTCCAGGCTTCCTGATCGG 666

Qy       667 AAAAGATTTGGGGCCAGGAACAATCAAAGAGAGCCTTCATTATGACTGA 711
             |||||||||||||||||||||||||||||||||||||||||||||||||
Db       667 AAAAGATTTGGGGCCAGGAACAATCAAAGAGAGCCTTCATTATGACTGA 711
```

RESULT 2
BX099226
LOCUS   BX099226   718 bp   mRNA   linear   EST 06-FEB-2003
DEFINITION   BX099226 Soares fetal liver spleen 1NFLS S1 Homo sapiens cDNA clone
              IMAGp998I061023 ; IMAGE:435149, mRNA sequence.
ACCESSION   BX099226
VERSION   BX099226.1 GI:27843984
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 718)
  AUTHORS   Ebert,L., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M.,
              Radelof,U., Schneider,D. and Korn,B.
  TITLE   Human UnigeneSet - RZPD3
  JOURNAL   Unpublished (2003)
COMMENT   Contact: Ina Rolfs
              RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
              Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
              RZPD; IMAGp998I061023.
              RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
              Human UnigeneSet - RZPD3 (RZPDLIB No.972)
              http://www.rzpd.de/CloneCards/cgi-

```
Db   541 TCTTCCAATGCTGTCCAGCTCAGTAGGTTTTTTGATAAGGTCATAGAGAAGAAGATACTTT 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   628 TTAAGAAGGTAATCAGATTCCAGGCTTTCCAGGCTTTCCTGATCGGAAAAAGATTGGGGCAGGAACA 687
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TTAAGAAGGTAATCAGATTCCAGGCTTTCCAGGCTTTCCTGATCGGAAAAAGATTGGGGCAGGAACA 660

Qy   688 TTAAAGAGAGCCTTCATTATGACTGA 711
             ||||||||||||||||||||||||||
Db   661 TTAAAGAGAGCCTTCATTATGACTGA 684
```

RESULT 3
BG501635
LOCUS   BG501635   720 bp   mRNA   linear   EST 27-MAR-2001
DEFINITION 602548678F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:4654986 5',
            mRNA sequence.
ACCESSION   BG501635
VERSION   BG501635.1 GI:13463152
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 720)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1443 row: k column: 19
            High quality sequence stop: 719.
FEATURES   Location/Qualifiers
     source   1..720
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4654986"
                     /tissue_type="embryonal carcinoma"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC 61"
                     /note="Organ: testis; Vector: pDNR-LIB (Clontech); Site_1:
                     Sfil (ggccgcctcggccc); Site_2: Sfil (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA. 5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C, G, or T). Average
                     insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
                     contained inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
                     Library."
ORIGIN
Query Match   95.9%; Score 682.2; DB 4; Length 720;
Best Local Similarity   99.3%; Pred. No. 5.5e-188;
Matches 706; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

```
Qy    1 ATGGGCTCCCTGGATCGGGTGAAGGTACTGGGTGTTGGGAGACTCAGGTGTTGGGAAATCT 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11 ATGGGCTCCCTGGATCGGGTGAAGGTACTGGGTGTTGGGAGACTCAGGTGTTGGGAAATCT 70

Qy   61 TCGTTAGTCAGTCCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATCATGGGGC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   71 TCGTTAGTCAGTCCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATCATGGGGC 130
```

```
Db   121 TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAAGAAGAGAGACCTAC 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   131 TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAAGAAGAGAGACCTAC 190

Db   181 TACATAGAATTATGGGATGTTGGAGGCTCGTGGGCCAGTGCCAGCAGCCTGAAAAGCACA 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   191 TACATAGAATTATGGGATGTTGGAGGCTCGTGGGCCAGTGCCAGCAGCCTGAAAAGCACA 250

Db   241 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTAACAAATAAG 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   251 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTAACAAATAAG 310

Db   301 AAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCA 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   311 AAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCA 370

Db   361 ACTGGAGTCTTGGTGACAAAATGGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATA 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   371 ACTGGAGTCTTGGTGACAAAATGGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATA 430

Db   421 CCACTGTTGGTAATAGGGACTAAACTGACCAGATTCATGAAACAAAGCGCCATGAAGTT 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   431 CCACTGTTGGTAATAGGGACTAAACCGGACCAGATTCATGAAACAAAGCGCCATGAAGTT 490

Db   481 TTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGC 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   491 TTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGC 550

Db   541 ACAAATCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCGTAGTAGGTTTT 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   551 ACAAATCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCGTCAGTAGG-TTTT 609

Db   601 GATAAGGTCATAGAGAAGGAGATACTTTTTAAGAGAAGAGATACTTTTTAAGAAGGTAATCAGATTCCAGGCTTTCC 659
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   610 GATAAGGTCATAGAGAAGGAGATACTTTTTAAGAGAAGAGATACTTTTTAAGAAGGTAATCAGATTCCAGGCTTTCC 669

Db   660 TGATCGGAAAAAGATTTGGGGCAGGAACAACATTAAAGAGAGCCTTCATTATGACTG 710
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   670 TGATCGGAAAAAGATTTGGGGCAGGAACAACATTAAAGAGAGCCTTCATTATGACTG 720
```

RESULT 4
BI333788
LOCUS   BI333788   732 bp   mRNA   linear   EST 30-JUL-2001
DEFINITION 602992795F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:5141563 5',
            mRNA sequence.
ACCESSION   BI333788
VERSION   BI333788.1 GI:15018445
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 732)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Incyte Genomics, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov.
            Plate: LLAM11348 row: m column: 20
            High quality sequence stop: 732.
FEATURES   Location/Qualifiers
     source   1..732
                     /organism="Homo sapiens"
                     /mol_type="mRNA"

REFERENCE
AUTHORS   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
           1 (bases 1 to 699)
           Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
           Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
           Lebkowski,J and Stanton,L.W.
TITLE   Transcriptome characterization elucidates signaling networks that
           control human ES cell growth and differentiation
JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT   Contact: Brandenberger R
           Regenerative Medicine
           Geron Corporation
           230 Constitution Drive, Menlo Park, CA 94025, USA
           Tel: 650 473 8658
           Fax: 650 473 7760
           Email: rbrandenberger@geron.com
           Insert Length: 699   Std Error: 0.00.
FEATURES   Location/Qualifiers
     source   1..699
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cells, DMSO-treated H9 cell
                     line"
                     /clone_lib="GRN PREHEP"
                     /note="oligo dT primed, full-length enriched cDNA library
                     from DMSO-treated hES cell line H9 (p22) maintained in
                     feeder-free conditions"
ORIGIN

Query Match   95.7%;   Score 680.6; DB 7;   Length 699;
Best Local Similarity 99.4%;   Pred. No. 1.6e-187;
Matches 683; Conservative   0; Mismatches   4; Indels   0; Gaps   0;

Qy   1 ATGGGCTCCCTGGATCGGGTGAAGTGTACTCGTGTTGGGAGACTCAGGTGTTGGGAAATCT 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13 ATGGGCTCCCTGGATCGGGTGAAGTGTACTCGTGTTGGGAGACTCAGGTGTTGGGAAATCT 72

Qy   61 TCGTTAGTCCATCTCCTATGCCAAAATCAAGTCTGGGAGACTGCTGGGCTGCTC 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73 TCGTTAGTCCATCTCCTATGCCAAAATCAAGTCTGGGAGACTGCTGGGCTGCTC 132

Qy   121 TGCTCAGTGATGTTCAGAGTTCATGATTACAAAGAGAGAAGAGACCTAC 180
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db   133 TGCTCAGTGATGTTCAGAGTTCATGATTACAAAGAGAGAAGAGACCTAC 192

Qy   181 TACATAGAATTATGGGACTGGATGTTGGAGGCTCTGTGGGCAGTGCTCTGCAGCAGCGTGAAAAGCACA 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   193 TACATAGAATTATGGGACTGGATGTTGGAGGCTCTGTGGGCAGTGCTCTGCAGCAGCGTGAAAAGCACA 252

Qy   241 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGTAACACTAACGAAATAAG 300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   253 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGTAACACTAACGAAATAAG 312

Qy   301 AAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATA 360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   313 AAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATA 372

Qy   361 ACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATA 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   373 ACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATA 432

Qy   421 CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTT 480
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   433 CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTT 492

Qy   481 TTAACTAGGACTGCTTTCCTGCGTCGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGC 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   493 TTAACTAGGACTGCTTTCCTGCGTCGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGC 552

Qy   541 ACAAATCCACGGTACTTAGCTGCAGGTTCTTCCCAATGCTGTCAAGCTCGTCAGTAGGTTTTTT 600
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   553 ACAAATCCACGGTACTTAGCTGCAGGTTCTTCCCAATGCTGTCAAGCTCGTCAGTAGGTTTTTT 612

/db xref="taxon:9606"
/clone="IMAGE:5141563"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/clone_lib="NIH_MGC_12"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.4 kb. Library prepared by Life
Technologies."
ORIGIN

Query Match   95.8%;   Score 681.4; DB 4;   Length 732;
Best Local Similarity 99.6%;   Pred. No. 9.4e-188;
Matches 704; Conservative   0; Mismatches   1; Indels   2; Gaps   2;

Qy   6 GTCCCTGGATCGGGTGAAGTGTACTCGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTT 65
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GTCCCTGGATCGGGTGAAGTGTACTCGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTT 60

Qy   66 AGTCCATCTCCTATGCCAAAATCAAGTCTGGGAGACTGCTGGGCTGCTC 125
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AGTCCATCTCCTATGCCAAAATCAAGTCTGGGAGACTGCTGGGCTGCTC 120

Qy   126 AGTGGATGTTCAGAGTTCATGATTACAAAGAGGAACCCCAGAAGAGAAGAGACCTACTACAT 185
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGTGGATGTTCAGAGTTCATGATTACAAAGAGGAACCCCAGAAGAGAAGAGACCTACTACAT 180

Qy   186 AGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGAAAAGCACAAGAGC 245
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGAAAAGCACAAGAGC 240

Qy   246 AGTATTCTACAACTCCGTAAATGTGTATTATTTTCGTACACGACGACTTAACAAATAAGAAGTC 305
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 AGTATTCTACAACTCCGTAAATGTGTATTATTTTCGTACACGACGACTTAACAAATAAGAAGTC 300

Qy   306 CTCCCAAAACTTGCGTCGTTGGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTTGGTGCCAACTGG 365
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CTCCCAAAACTTGCGTCGTTGGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTTGGTGCCAACTGG 360

Qy   366 AGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATACCACT 425
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AGTCTTGGTTGCAAATGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATACCACT 420

Qy   426 GTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTTTTAAC 485
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTTTTAAC 480

Qy   486 TAGGACTGCTTTCCTGGCTGAGGATTCAATCCAGAAGAAATTAATTTGGACTGCACAAA 545
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TAGGACTGCTTTCCTGGCTGAGGATTCAATCCAGAAGAAATTAATTTGGACTGCACAAA 540

Qy   546 TCCACGGTACTTAGCTGCAGGTTCTTCCCAATGCTGTCAAGCTCGTCAGTAGGTTTTTTGATAA 605
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TCCACGGTACTTAGCTGCAGGTTCTTCCCAATGCTGTCAAGCTCGTCAGTAGGTTTTTTGATAA 600

Qy   606 GGTCATAGAGAAGAGATACTTTT-AAGAGAAGTAATCAGATTCAGGCTTTCCAGGCTTTCCTGATC 664
            ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   601 GGTCATAGAGAAGAGATACTTTTTTAAGAGAAGTAATCAGATTCAGGCTTTCCAGGCTTTCCTGATC 660

Qy   665 GGAAAAGAGATTTGGGGCAGGAACATTAAAGAGAGCCCTTCATTATGACTGA 711
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 GGAAAAGA-TTGGGGCAGGAACATTAAAGAGAGCCCTTCATTATGACTGA 706

RESULT 5
CN361072
LOCUS   CN361072   699 bp   mRNA   linear   EST 16-MAY-2004
DEFINITION 17000600016086 GRN_PREHEP Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN361072
VERSION   CN361072.1 GI:47361006
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens

```
Qy  331 TTGGAAGCTCTCAACACGGGATTTGGTGCCAACTGGACTTCGGTGAGTCTTGGTGACAAATGGGGATTAT 390
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 TTGGAAGCTCTCAACACGGGATTTGGTGCCAACTGGACTTCGGTGAGTCTTGGTGACAAATGGGGATTAT 360

Qy  391 GATCAAGAACAGTTGTCGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGAC 450
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 GATCAAGAACAGTTGTCGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGAC 420

Qy  451 CAGATTCATGAAACAAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGGAT 510
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 CAGATTCATGAAACAAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGGAT 480

Qy  511 TTCAATCCAGAAGAAAATTAATTTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCT 570
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 TTCAATCCAGAAGAAAATTAATTTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCT 540

Qy  571 TCCAATGCTGTCAAGCTCAGTAGGTTTTTGATAAGGTCATAGAAGAAGAGATACTTTTTA 630
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 TCCAATGCTGTCAAGCTCAGTAGGTTTTTTGATAAGGTCATAGAAGAAGAGATACTTTTTA 600

Qy  631 AGAGAAGGTAATCAGATTCAGATTCCAGGCTTTCCTCGTGATCGGAAAAGATTTGGGGCAGGAACATTA 690
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 AGAGAAGGTAATCAGATTCAGATTCCAGGCTTTCCTCGTGATCGGAAAAGATTTGGGGCAGGAACATTA 660

Qy  691 AAGAGCCTTCATTATGACTGA 711
            |||||||||||||||||||||
Db  661 AAGAGCCCTTCATTATGACTGA 681
```

RESULT 7
BQ215951
LOCUS   BQ215951   815 bp   mRNA   linear   EST 02-MAY-2002
DEFINITION AGENCOURT_7558711 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6045318
            5', mRNA sequence.
ACCESSION   BQ215951
VERSION   BQ215951.1 GI:20397351
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 815)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
  TITLE   Unpublished (1999)
  JOURNAL   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
COMMENT   Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13289 row: f column: 07
            High quality sequence stop: 614.
FEATURES   Location/Qualifiers
     source   1..815
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6045318"
                     /tissue_type="embryonal carcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_92"
                     /note="Organ: testiS; Vector: pCMV-SPORT6; Site 1: NotI;
                     Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 2.5 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
ORIGIN

Query Match   94.9%; Score 674.8; DB 5; Length 815;

```
Qy  601 GATAAGGTCATAGAAGAAGAGATACTTTTTAAGAGAAGGTAATCAGATTCCAGGCTTTCCT 660
            ||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db  613 GATAAGGTCATAGAAGAAGAGATACTTTTTAAGAGAAGGTAATCAGATTCCAGGCTTTCCT 672

Qy  661 GATCGGAAAAGATTTGGGGCAGGAACA 687
            ||||||| || |||||||||||||||
Db  673 GATCGGAAAAGATTTGGGGCAGGAACA 699
```

RESULT 6
CN361073
LOCUS   CN361073   720 bp   mRNA   linear   EST 16-MAY-2004
DEFINITION 32877509 GRN_ES Homo sapiens cDNA 5', mRNA sequence.
ACCESSION   CN361073
VERSION   CN361073.1 GI:47361007
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 720)
  AUTHORS   Brandenberger,R., Wei,H., Zhang,S., Lei,S., Murage,J., Fisk,G.J.,
            Li,Y., Xu,C., Fang,R., Guegler,K., Rao,M.S., Mandalam,R.,
            Lebkowski,J and Stanton,L.W.
  TITLE   Transcriptome characterization elucidates signaling networks that
            control human ES cell growth and differentiation
  JOURNAL   Nat. Biotechnol. 22 (6), 707-716 (2004)
COMMENT   Contact: Brandenberger R
            Regenerative Medicine
            Geron Corporation
            230 Constitution Drive, Menlo Park, CA 94025, USA
            Tel: 650 473 8658
            Fax: 650 473 7760
            Email: rbrandenberger@geron.com
            Insert Length: 720   Std Error: 0.00.
FEATURES   Location/Qualifiers
     source   1..720
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="embryonic stem cells, cell lines H1, H7, and
                     H9"
                     /clone_lib="GRN_ES"
                     /note="oligo dT primed, full-length enriched cDNA library
                     from undifferentiated hES cell lines H1 (p32), H7 (p29),
                     and H9 (p26) maintained in feeder-free conditions",
ORIGIN

Query Match   95.6%; Score 679.4; DB 7; Length 720;
Best Local Similarity   99.9%;
Matches   680; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

```
Qy   31 GTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAA 90
            | |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 GGGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAA 60

Qy   91 GTGCTCGGGAAATCCATCATGACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTAC 150
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GTGCTCGGGAAATCCATCATGACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTAC 120

Qy  151 AAAGAAGGAACCCCAGAAGAGAAGAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCT 210
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 AAAGAAGGAACCCCAGAAGAGAAGAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCT 180

Qy  211 GTGGGCAGTGCCAGCAGCCGTGAAAAGCAACAAGAGCAGTATTCTACAACTCCGTAAATGGT 270
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GTGGGCAGTGCCAGCAGCCGTGAAAAGCAACAAGAGCAGTATTCTACAACTCCGTAAATGGT 240

Qy  271 ATTATTTTCGTACACGACTTAACAAATAAGAAGTCCTCCCAAAAACTTGCCGTCGTTGGTCA 330
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 ATTATTTTCGTACACGACTTAACAAATAAGAAGTCCTCCCAAAAACTTGCCGTCGTTGGTCA 300
```

Best Local Similarity   99.6%;   Pred. No. 8.2e-186;
Matches 676;   Conservative   0;   Mismatches   3;   Indels   0;   Gaps   0;

Qy   33 GTTGGGAGACTCAGTGTGTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAATCAAGT   92
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GTTGGGAGACTCAGTGTGTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAATCAAGT   60

Qy   93 GCTGGGAAATCCATCATGGACTGTGGGCTGCTCAGTGATGTCAGAGTTCATGATTACAA 152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GCTGGGAAATCCATCATGGACTGTGGGCTGCTCAGTGATGTCAGAGTTCATGATTACAA 120

Qy   153 AGAAGGAACCCAGAAGAGAAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGT 212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 AGAAGGAACCCAGAAGAGAAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGT 180

Qy   213 GGGCAGTGCCAGCAGCGTGAAAAGCACAAGAGCAAGAGCAGTATTCTACAACTCCGTAAATGGTAT 272
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GGGCAGTGCCAGCAGCGTGAAAAGCACAAGAGCAAGAGCAGTATTCTACAACTCCGTAAATGGTAT 240

Qy   273 TATTTTCGTTACACGACTTAACAAATAAGAAGTCCTCCCAAAAACTTGCGGTCGTTGGTCATT 332
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TATTTTCGTTACACGACTTAACAAATAAGAAGTCCTCCCAAAAACTTGCGGTCGTTGGTCATT 300

Qy   333 GGAAGCTCTCAACAGGGATTGGTGCCAACTGGAGTCTTGGTGACAAATGGGGATTATGA 392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GGAAGCTCTCAACAGGGATTGGTGCCAACTGGAGTCTTGGTGACAAATGGGGATTATGA 360

Qy   393 TCAAGAACAGTTTGCTGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCA 452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TCAAGAACAGTTTGCTGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCA 420

Qy   453 GATTCATGAAACAAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGGGATTT 512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 GATTCATGAAACAAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGGGATTT 480

Qy   513 CAATCCAGAAGAGAAATTAATTTGGACTGCGCACAAATCCACGGTACTTAGCTGTCCTTC 572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CAATCCAGAAGAGAAATTAATTTGGACTGCGCACAAATCCACGGTACTTAGCTGTCCTTC 540

Qy   573 CAATGCTGTCAAGCTCAGTAGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTTAAG 632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CAATGCTGTCAAGCTCAGTAGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTTAAG 600

Qy   633 AGAAGGTAATCAGATTCCAGGCTTCCAGGCTTCCTGATCGGAAAAGATTTGGGGCAGGAACATTAAA 692
          ||||||||||||||||||||||||||||||||| |||||||||||||||
Db   601 AGAAGGTAATCAGATTCCAGGCTTCCAGGCTTCCTGATNCGGAAAAGATTTGGGGCAGGAACATTAAA 660

Qy   693 GAGCCTTCATTATGACTGA 711
          |||||||||||||||||||
Db   661 GAGCCTTCATTATGACTGA 679

RESULT 8

CR619977

LOCUS   CR619977   1867 bp   mRNA   linear   HTC 21-JUL-2004
DEFINITION   full-length cDNA clone CS0DA005YF10 of Neuroblastoma of Homo
             sapiens (human).
ACCESSION   CR619977
VERSION   CR619977.1 GI:50500784
KEYWORDS   HTC; CNSLT_cDNA.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1867)
AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE   Full-length cDNA libraries and normalization
JOURNAL   Unpublished
REMARK   Contact: Feng Liang Email : fliang@lifetech.com URL :
             http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
             Faraday Avenue
REFERENCE   2 (bases 1 to 1867)
AUTHORS   Genoscope.

TITLE   Direct Submission
JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
             BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
             - Web : www.genoscope.cns.fr)

COMMENT   1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
             end enriched, double-strand cDNA was digested with Not I and cloned
             into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
             was normalized. Library was constructed by Life Technologies, a
             division of Invitrogen.

FEATURES   Location/Qualifiers
     source   1..1867
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DA005YF10"
                     /tissue_type="Neuroblastoma"
                     /plasmid="pCMVSPORT_6"
ORIGIN

Query Match   94.0%;   Score 668;   DB 3;   Length 1867;
Best Local Similarity 100.0%;   Pred. No. 1e-183;
Matches 668;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   44 CAGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATC 103
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 CAGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATC   60

Qy   104 CATCATGGACTGTGGGCTGCTCAGTGATGTCAGAGTTCATGATTACAAAGAAGGAACCC 163
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CATCATGGACTGTGGGCTGCTCAGTGATGTCAGAGTTCATGATTACAAAGAAGGAACCC 120

Qy   164 CAGAAGAGAAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCA 223
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 CAGAAGAGAAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCA 180

Qy   224 GCAGCGTGAAAAGCACAAGAGCAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTAC 283
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GCAGCGTGAAAAGCACAAGAGCAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTAC 240

Qy   284 ACGACTTAACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCA 343
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ACGACTTAACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCA 300

Qy   344 ACAGGGATTGGTGCCAACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGAACAGT 403
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ACAGGGATTGGTGCCAACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGAACAGT 360

Qy   404 TTGCTGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAA 463
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 TTGCTGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAA 420

Qy   464 CAAAGGCCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAG 523
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CAAAGGCCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAG 480

Qy   524 AAATTAATTTGGACTGCGCACAAATCCACGGTACTTAGCTGTCCCAATGCTGTCA 583
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 AAATTAATTTGGACTGCGCACAAATCCACGGTACTTAGCTGTCCCAATGCTGTCA 540

Qy   584 AGCTCAGTAGGTTTTTTGATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAATC 643
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 AGCTCAGTAGGTTTTTTGATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAATC 600

Qy   644 AGATTCCAGGCTTCCTGATCGGAAAAGATTTGGGGCAGGAACATTAAAAGAGCCTTCATT 703
          ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   601 AGATTCCAGGCTTCCTGATCGGAAAAGATTTGGGGCAGGAACATTAAAAGAGCCTTCATT 660

Qy   704 ATGACTGA 711
          ||||||||
Db   661 ATGACTGA 668

RESULT 9

```
AL516309
LOCUS       AL516309     797 bp    mRNA    linear   EST 23-MAR-2004
DEFINITION  AL516309 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
            CS0DA005YF10 5-PRIME, mRNA sequence.
ACCESSION   AL516309
VERSION     AL516309.3  GI:45552953
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 797)
AUTHORS     Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE       Full-length cDNA libraries and normalization
JOURNAL     Unpublished (2001)
COMMENT     On Feb 13, 2001 this sequence version replaced gi:30490812.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            2 rue Gaston Cremieux, CP 5706 - 91057 EVRY cedex - FRANCE
            Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
            1st strand cDNA was primed with a Not-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library
            was not normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
            This sequence belongs to sequence cluster 1912.r,
            and it belongs to a clone representative of this cluster. For more
            information about this cluster and the virtual cDNA, see
            http://www.genoscope.cns.fr/cdna?s=CS0DA005DC05QP1&c=1912.r.

FEATURES    Location/Qualifiers
   source   1..797
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0DA005YF10"
            /tissue_type="NEUROBLASTOMA"
            /clone_lib="Homo sapiens NEUROBLASTOMA"
            /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
            with a NotI-oligo(dT) primer. Five prime end enriched,
            double-strand cDNA was digested with Not I and cloned into
            the Not I and EcoRV sites of the pCMVSPORT 6 vector.
            Library was not normalized."

ORIGIN

Query Match           93.3%;   Score 663.6;   DB 1;   Length 797;
Best Local Similarity 98.8%;   Pred. No. 1.5e-182;
Matches 660;   Conservative 7;   Mismatches 1;   Indels 0;   Gaps 0;

Qy   44 CAGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATC 103
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 CAGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATC  60

Qy  104 CATCATGGACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCC 163
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 CATCATGGACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCC 120

Qy  164 CAGAAGAGAAGACCTACTACATAGATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCA 223
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CAGAAGAGAAGACCTACTACATAGATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCA 180

Qy  224 GCAGGCGTGAAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTAC 283
         |||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GCCGGCGTGAAAAGCACAAGAGCAGTATTCTCACAACTCCGTAAATGGTATTATTTTCGTAC 240

Qy  284 ACGACTTAACAAATAAGAAGTCCTCCCAAAACTTCGGTCGTTCATTGGAAGCTCTCA 343
         |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db  241 ACGACTTAACAAATAAGAAGTCCTCCCAAAACTTCGGTCGTTCATTGGAAGCYCTCA 300

Qy  344 ACAGGGATTTGGTGCCAACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGAGACAGT 403
         ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db  301 ACAGGGATTTGGTGCCAACTGGAGTCTTGGTGACAAATGGGGATTATGATCAMGAAGACAGT 360

Qy  404 TTGCTGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAA 463
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 TTGCTGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCAATCAGAAG 420

Qy  464 CAAAGCGCCATGAAGTTTAACTAGGACTGCYTTCCTGGCTGAGGATTCAATCCAGAAG 523
         ||||||| |||||||||||||||||||||| ||||||||||||||||||||||||||
Db  421 CAAAGCGCATGAAGTTTTAACTAGGACTGCYYTCCTGGCTGAGGATTCAATCCAGAAG 480

Qy  524 AAATTAATTTGGACTGCACAAATCCACGGTACTTAGCTGCTCAGTTGTTCTTCCAATGCTGTCA 583
         |:|||||||||||||||||||||||||||||||||||||||||:||||||||||||||:||||
Db  481 AMATTAATTTGGACTGCACAAATCCACGGTACTTAGCTGCMGGTTCTTCCAATGCYGTCA 540

Qy  584 AGCTCAGTAGTTTTTGATAAGTCATAGAAGGTCATAGAGAAGAGATACTTTTAAGAGAAGAAGGTAATC 643
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 AGCTCAGTAGTTTTTGATAAGGTCATAGAAGGTCATAGAGAAGAGATACTTTTAAGAGAAGAAGGTAATC 600

Qy  644 AGATTCCAGGCTTTCCTGATCCGAAAAGATTTGGGCAGGAACATTAAAGAGCCTTCATT 703
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 AGATTCCAGGCTTTCCTGATCCGAAAAGATTTGGGCAGGAACATTAAAGAGCCTTCATT 660

Qy  704 ATGACTGA 711
         ||||||||
Db  661 ATGACTGA 668


RESULT 10
BM451532
LOCUS       BM451532     989 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT 6395013 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5493017
            5', mRNA sequence.
ACCESSION   BM451532
VERSION     BM451532.1  GI:18500572
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 989)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12116 row: a  column: 18
            High quality sequence stop: 652.

FEATURES    Location/Qualifiers
   source   1..989
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:5493017"
            /tissue_type="retinoblastoma"
            /lab_host="DH10B (phage-resistant)"
            /clone_lib="NIH MGC 67"
            /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
            Average insert size 1.75 kb. Library constructed by Life
            Technologies."

ORIGIN

Query Match           92.0%;   Score 654.4;   DB 4;   Length 989;
Best Local Similarity 99.7%;   Pred. No. 8e-180;
Matches 666;   Conservative 0;   Mismatches 1;   Indels 1;   Gaps 1;

Qy   45 AGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCC 104
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  182 AGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCC 241
```

```
FEATURES    Location/Qualifiers
    source    1..902
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="IMAGE:4510560"
        /tissue_type="embryonal carcinoma, cell line"
        /lab_host="DH10B (phage-resistant)"
        /clone_lib="NIH_MGC_92"
        /note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
        Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
        Average insert size 2.5 kb. Library enriched for
        full-length clones and constructed by Life Technologies.
        Note: this is a NIH_MGC Library."

ORIGIN

Query Match    91.4%;    Score 650.2;    DB 4;    Length 902;
Best Local Similarity    98.7%;    Pred. No. 1.3e-178;
Matches 666;    Conservative 0;    Mismatches 8;    Indels 1;    Gaps 1;

Qy    19 GTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTAGTCCATCTCCTA 78
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 GTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTAGTCCATCTCCTA 60

Qy    79 TGCCAAAATCAAGTGCTGGGAAATCATCATGACTGTGGGCTGCTCAGTGATGTGATCAGA 138
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TGCCAAAATCAAGTGCTGGGAAATCATCATGACTGTGGGCTGCTCAGTGATGTGATCAGA 120

Qy   139 GTTCATGATTACAAAGAAGAACCCAGAGAGAGAAGACCTACTACATAGAATTATGGGAT 198
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GTTCATGATTACAAAGAAGAACCCAGAGAGAGAAGACCTACTACATAGAATTATGGGAT 180

Qy   199 GTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGAAAAGCACAAGAGCAGTATTCTACAAC 258
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGAAAAGCACAAGAGCAGTATTCTACAAC 240

Qy   259 TCCGTAAATGTATTATTTCGTACACGACTTAACAAATAAGAAGTCCTCCCAAAACTTG 318
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TCCGTAAATGTATTATTTCGTACACGACTTAACAAATAAGAAGTCCTCCCAAAACTTG 300

Qy   319 CGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGCTGATAACCAAATACCACTGTTGGTGACA 378
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGCTGATAACCAAATACCACTGTTGGTGACA 360

Qy   379 AATGGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATACCACTGTTGCTGGTAATAGGG 438
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AATGGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATACCACTGTTGCTGGTAATAGGG 420

Qy   439 ACTAAACTGGACCCAGATTCATTGAAACAAAGGCCATGAAGTTTAACTAGGACTGCTTTC 498
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ACTAAACTGGACCCAGATTCATTGAAACAAAGGCCATGAAGTTTAACTAGGACTGCTTTC 480

Qy   499 CTGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGCACAAATCCACGGTACTTA 558
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 CTGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGCACAAATCCACGGTACTTA 540

Qy   559 GCTGCAGGTTCTTCCAATGCTGTCCAAGCTCAGTAGTTTTTTGATAAGGTCATAGAGAAG 618
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 GCTGCAGGTTCTTCCAATGCTGTCCAAGCTCAGTAGTTTTTTCGATAAGGTCATAGAGAAG 600

Qy   619 AGATACTTTTAAGACAGAAGGTAATCAGATTCCAGGCTTCCTGATCCGGAA-AAGATTGG 677
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 AGATACTTTCTAAGACAGAAGGTAATCAGATTCCAGGTTTTCCTGATCCGGCGTTTTCCGC 660

Qy   678 GGCCAGGAACATTAAA 692
         ||||| ||||
Db   661 GGCCAGGAACATATAA 675
```

```
    105 ATCATGGACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAACCCC 164
Qy      |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  242 ATCATGGACTGTGGGCTGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAACCCC 301

    165 AGAAGAAGAAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAG 224
Qy      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  302 AGAAGAAGAAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAG 361

    225 CAGCGTGAAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGTATTATTTCGTACA 284
Qy      ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  362 CAGCGTGAAAAGCACAAGAGCAGTATTCTACAACTCCGTAAATGTATTATTTCGTACA 421

    285 CGACTTAACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAA 344
Qy      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  422 CGACTTAACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAA 481

    345 CAGGGATTTGGTGCGCCAACTGGAGTCTTGGTGACAAAATGGGGATTATGATCAAGAACAGTT 404
Qy      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  482 CAGGGATTTGGTGCGCCAACTGGAGTCTTGGTGACAAAATGGGGATTATGATCAAGAACAGTT 541

    405 TGCTGATAACCACCACTGTTGGTAATAGGACTAAACTGGACCCAGATTCATTGAAAC 464
Qy      |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  542 TGCTGATAACCACCACTGTTGGTAATAGGACTAAACTGGACCCAGATTCATTGAAAC 601

    465 AAAGCGCCATGAAGTTTAAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGA 524
Qy      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  602 AAAGCGCCATGAAGTTTAAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGA 661

    525 AATTAATTTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAA 584
Qy      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  662 AATTAATTTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAA 721

    585 GCTCAGTAGTTTTTTGATAAGGTCATAGAGAGAAGAGATACTTTTTGATAAGGTAATCA 644
Qy      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  722 GCTCAGTAGTTTTTTGATAAGGTCATAGAGAGAAGAGATACTTTTTGATAAGGTAATCA 781

    645 GATTCCAGGC-TTTCCTGATCCGGAAAAAGATTTGGGGCAGGAACATTAAAGAGCCTTCATT 703
Qy      ||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db  782 GATTCCAGGCGTTTCCTGATCCGGGGAAGATTTGGGGCAGGAACATTAAAGAGCCCTTCATT 841

    704 ATGACTGA 711
Qy      ||||||||
Db  842 ATGACTGA 849
```

RESULT 11
BG258348
LOCUS   BG258348   902 bp   mRNA   linear   EST 13-FEB-2001
DEFINITION 602379905F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4510560 5',
            mRNA sequence.
ACCESSION   BG258348
VERSION   BG258348.1   GI:12768164
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 902)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10392 row: j column: 01
            High quality sequence stop: 648.

RESULT 12
CK455329
LOCUS   CK455329   855 bp   mRNA   linear   EST 12-JAN-2004
DEFINITION 915906 MARC 4PIG Sus scrofa cDNA 5', mRNA sequence.

ACCESSION   CK455329
VERSION   CK455329.1 GI:40802543
KEYWORDS   EST.
SOURCE   Sus scrofa (pig)
  ORGANISM. Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1 (bases 1 to 855)
  AUTHORS   Smith,T.P.L., Freking,B.A., Ford,J.J., Vallet,J.L., Wise,T.A.,
            Nonneman,D.J., Wray,J.E. and Keele,J.W.
  TITLE   Porcine EST collection using a normalized library constructed from
            embryos representing early developmental stages
  JOURNAL   Unpublished (2003)
  COMMENT   Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called with phred v0.020425.c and
            trimmed with the aid of the trim_alt option. Vector identified with
            cross_match v0.990329.
            Plate: TMW8028 row: B column: 12
            Seq primer: GTAATACGACTCACTATAGGG.
FEATURES   Location/Qualifiers
     source 1..855
             /organism="Sus scrofa"
             /mol_type="mRNA"
             /db_xref="taxon:9823"
             /tissue_type="pooled"
             /lab_host="DH10B"
             /clone_lib="MARC 4PIG"
             /note="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI;
             Library made with combined RNA from day-10, day-13,
             day-15, day-25, and day-30 whole embryos."
ORIGIN

Query Match   91.4%; Score 650; DB 7; Length 855;
Best Local Similarity   95.0%; Pred. No. 1.5e-178;
Matches   671; Conservative   0; Mismatches   35; Indels   0; Gaps   0;

Qy   6 GTCCCTGGATCGGGTGAAGGTACTCGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTT 65
            ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   1 GTCCCTGGATCGGGTGAAAGTACTCGTGTGTTGGGAGACTCAGGTGTGTTGGGAAATCTTCACT 60

Qy   66 AGTCCATCTCCTATGCCAAAATCAAGTGCTCGGGAAATCCATCATGGACTCGTGGGCTGCTC 125
            ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AGTTCATCTTCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACTCGTGGGCTGCTC 120

Qy   126 AGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAGAAGAAGACCTACTACAT 185
            ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db   121 AGTAGATGTCAGAGTTCATGATTCATGACTACACAAAGAAGGAACCCCAGAAGAGAAGAAGACCTACTATAT 180

Qy   186 AGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCGCGTGAAAGCACAAGAGC 245
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 AGAATTATGGGATGTTGGAGGCTCGTGCGTGGGCAGTGCCAGCAGTGTGAAAAGCACAAGAGC 240

Qy   246 AGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTAACAAATAAGAAGTC 305
            ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   241 AGTATTCTACAACTCCGTAAATGGTATTATTTTAGTACACGACTTAACATGTAACAAATAAGAAGTC 300

Qy   306 CTCCAAAACTTGCCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCAACTGG 365
            ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ATCCCAAAATTTGTATCGTTGGTCATTGGAAGCTCTCAACAGGGGATCTCGGTGCCAACTGG 360

Qy   366 AGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTGCTGATAACCAAATACCACT 425
            |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db   361 AGTCTTGGTAACAAATGGGATTATGACCGGGAACAGTTTGCTGATAACCAAATCCCACT 420

Qy   426 GTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTTTAAC 485
            ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   421 GTTGGTAATAGGGACTAAACTGGATCAGATTCAGATTCATGAAACCAAGCGCCATGAAGTTTAAC 480

Qy   3 GGCGTCCCTGGATCGGGTGAAGGTACTCGTGTTGGGAGACTCAGGTGTTGGGAAATCTT- 61
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGCGTCTCGGGATCGGGTGAAGGTACTCGTGTTGGGAGACTCAGGTGTTGGGAAATCTTG 60

Qy   486 TAGGACTGCTTCCTGCTGAGGATTCCAATCCAGAGAAATTAATTTGGACTGCACAAA 545
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 TAGGACTGCTTTCCTGCTGAGGATTTCAATCCAATGCTTCAAGCTCAGTAGGTTTTTGATAA 540

Qy   546 TCCACGGTACTTAGCTGCAGTTCTTCCAATGCTTCAAGCTCAGTAGGTTTTTTGATAA 605
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 TCCACGGTACTTAGCTGCAGTTCTTCCAATGCTGTCAAGCTCAGTAGGTTTTTTGATAA 600

Qy   606 GGTCATAGAGAGAGAGATACTTTTAAGAGAAGAAGTAATCAGATTCCAGGCTTCCTGATCG 665
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 GGTCATAGAGAAGAAGATACTTTTTAAGAGATGATGTAATCAGATTCCAGGCTTCCTGATCG 660

Qy   666 GAAAAGATTTGGGCAGGAACATTAAAGAGCCTTCATTATGACTGA 711
            |||||||||||||||||||||||||||||||||||||||||||||
Db   661 AAAAAGGTTTGGAGGAGGAACATTAAAGAGCTTACATTATGACTGA 706

RESULT 13
BI759699
LOCUS   BI759699   785 bp   mRNA   linear   EST 25-SEP-2001
DEFINITION 603045561F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5186006 5',
            mRNA sequence.
ACCESSION   BI759699
VERSION   BI759699.1 GI:15751277
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 785)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11464 row: i column: 15
            High quality sequence start: 13
            High quality sequence stop: 761.
FEATURES   Location/Qualifiers
     source 1..785
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /clone="IMAGE:5186006"
             /lab_host="DH10B"
             /clone_lib="NIH_MGC_116"
             /note="Organ: pooled colon, kidney, stomach; Vector:
             pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
             source anonymous pool of 3 colons, age 26 yo male, 49 yo
             female, 71 yo male colon; 46 yo male kidney, and pool of 2
             stomachs, 62 yo male and 70 yo female. Library is
             oligo-dT primed and directionally cloned (EcoRV site is
             destroyed upon cloning). Average insert size 1.4 kb,
             insert size range 1-3 kb. Library is normalized and
             enriched for full-length clones and was constructed by C.
             Gruber (Invitrogen). Research Genetics tracking code
             023. Note: this is a NIH_MGC Library."
ORIGIN

Query Match   90.6%; Score 644; DB 4; Length 785;
Best Local Similarity   97.8%; Pred. No. 8.3e-177;
Matches   696; Conservative   0; Mismatches   10; Indels   6; Gaps   4;

Qy   3 GGCGTCCCTGGATCGGGTGAAGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTT- 61
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGCGTCTCGGGATCGGGTGAAGGTACTCGTGTTGGGAGACTCAGGTGTTGGGAAATCTTG 60

FEATURES   Location/Qualifiers
    source   1..649
         /organism="Homo sapiens"
         /mol_type="mRNA"
         /db_xref="taxon:9606"
         /clone="IMAGE:4362643"
         /tissue_type="lymphoma, cell line"
         /lab_host="DH10B (phage-resistant)"
         /clone_lib="NIH_MGC_85"
         /note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
         Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
         Average insert size 1.867 kb. Library enriched for
         full-length clones and constructed by Life Technologies.
         Note: this is a NIH_MGC Library."

ORIGIN

Query Match   89.5%;   Score 636.4;   DB 4;   Length 649;
Best Local Similarity   99.7%;
Matches 648;   Conservative   0;   Mismatches   1;   Indels   1;   Gaps   1;

```
Qy   28  CTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTTAGTCCATCTCCTATGCCAAAAT  87
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  CCGGTGTTGGGAGACTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGG  120

Qy   88  CAAGTGCTGGGAAATCCATCATGGACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGAT  147
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  CAAGTGCTGGGAAATCCATCATGGACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGAT  120

Qy  148  TACAAGAAGGAACCCCAGAAGAGAAGACCTACTACATAGAATTATGGGATGTTGGAGGC  207
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  TACAAAGAAGGAACCCCAGAAGAGAAGACCTACTACATAGAATTATGGGATGTTGGAGGC  180

Qy  208  TCTGTGGGCAGTGCCAGCAGCGTGAAAAGCACAAGAGCAGTATTCTACAACTCCGTAAAT  267
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TCTGTGGGCAGTGCCAGCAGCGTGAAAAGCACAAGAGCAGTATTCTACAACTCCGTAAAT  240

Qy  268  GGTATTATTTTCGTACACGACTTAACAAATAAGAAGTCCTCCCAAAACTTCGGTCGTTGG  327
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  GGTATTATTTTCGTACACGACTTAACAAATAAGAAGTCCTCCCAAAACTTCGGTCGTTGG  300

Qy  328  TCATTGGAAGCTCTCAACACAGGGATTGGTGCCAACTGTCTTGGTGACAAAATGGGGGAT  387
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  TCATTGGAAGCTCTCAACACAGGGATTGGTGCCAACTGTCTTGGTGACAAAATGGGGGAT  360

Qy  388  TATGATCAAGAACAGTTTGCTGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTG  447
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  TATGATCAAGAACAGTTTGCTGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTG  420

Qy  448  GACCAGATTCATGAAACAAAGCGCCATGAAGTTTAACTAGACTGCTTTCCTGGCTGAG  507
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  GACCAGATTCATGAAACAAAGCGCCATGAAGTTTAACTAGACTGCTTTCCTGGCTGAG  480

Qy  508  GATTTCAATCCAGAGAAAATTAATTTGGACTGCACAAATCCACGGTACTTAGCTGCAGGT  567
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  GATTTCAATCCAGAGAAAATTAATTTGGACTGCACAAATCCACGGTACTTAGCTGCAGGT  540

Qy  568  TCTTCCAATGCTGTCAAGCTCAGTAGTTTTTTGATAAGGTCATAGAAGAAGAGATACTTT  627
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  TCTTCCAATGCTGTCAAGCTCAGTAGG-TTTTGATAAGGTCATAGAAGAAGAGATACTTT  599

Qy  628  TTAAGAAGTAATCAGATTCCAGGCTTTCCTGATCGGAAAAGATTTGG  677
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db  600  TTAAGAGAAGTACTCAGATTCCTCCAGGCTTTCCTGATCGGAAAAGATTTGG  649
```

RESULT 15
BG031602
LOCUS   BG031602   959 bp   mRNA   linear   EST 24-JAN-2001
DEFINITION   602299747F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4394290 5',
         mRNA sequence.
ACCESSION   BG031602
VERSION   BG031602.1   GI:12422045
KEYWORDS   EST.

```
Qy   62  --CGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGG  119
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  CCGGTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGG  120

Qy  120  CTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAGAAGACCTA  179
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  CTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAGAAGACCTA  180

Qy  180  CTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCGGTGAAAAGCAC  239
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  CTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCGGTGAAAAGCAC  240

Qy  240  AAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTAACAAATAA  299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  AAGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTAACAAATAA  300

Qy  300  GAAGTCCTCCCAAAACTTCGGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTGGTGCC  359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  GAAGTCCTCCCAAAACTTCGGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTGGTGCC  360

Qy  360  AACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGAACAGTTTGCTGATAACCAAAT  419
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  AACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGAACAGTTTGCTGATAACCAAAT  420

Qy  420  ACCACTGTTGGTAATAGGGACTAAACTGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGT  479
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  ACCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGT  480

Qy  480  TTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAGAAATTAATTTGGACTG  539
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  TTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAGAAATTAATTTGGACTG  540

Qy  540  CACAAATCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCGTGTAGTTTT  599
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  CACAAATCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCTCGTGTAGTAGG-TTTT  599

Qy  600  TGATAAGGTCATAGAAGAAGAGATACTTTTTAAGAAGTAATCAGATTCCAGGCTTTCC  659
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  600  TGATAAGGTCATAGAAGAAGAGATAC-TTTTAAGAAGTAATCAGATTCATTCCAGGCTTTCC  658

Qy  660  TGATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGCCTTCATTATGACTGA  711
             ||||||| || ||||||||||||||||||||||||||||||||||
Db  659  TGATCGGAACAAGATTTGGGCAGGAACATT-AAGAGCCTTCATTATGACTGA  709
```

RESULT 14
BG025782
LOCUS   BG025782   649 bp   mRNA   linear   EST 24-JAN-2001
DEFINITION   602274775F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4362643 5',
         mRNA sequence.
ACCESSION   BG025782
VERSION   BG025782.1   GI:12412725
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
         Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
         Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 649)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
  COMMENT   Contact: Robert Strausberg, Ph.D.
         Email: cgapbs-r@mail.nih.gov
         Tissue Procurement: Louis Staudt, M.D., Ph.D.
         cDNA Library Preparation: Life Technologies, Inc.
         cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
         DNA Sequencing by: Incyte Genomics, Inc.
         Clone distribution: MGC clone distribution information can be
         found through the I.M.A.G.E. Consortium/LLNL at:
         http://image.llnl.gov
         Plate: LLAM10007   row: f   column: 20
         High quality sequence stop: 649.

```
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 959)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM10089  row: m  column: 11
             High quality sequence stop: 665.
FEATURES             Location/Qualifiers
     source          1..959
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4394290"
                     /tissue_type="mammary adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_87"
                     /note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
                     Average insert size 1.383 kb. Library enriched for
                     full-length clones and constructed by Life Technologies.
                     Note: this is a NIH_MGC Library."
ORIGIN

Query Match         88.3%;  Score 628;  DB 4;  Length 959;
Best Local Similarity   99.4%;  Pred. No. 4.2e-172;
Matches  662;  Conservative   0;  Mismatches   0;  Indels   4;  Gaps   3;

Qy         48 TGTTGGGAAATCTTCGTTAGTCCATCTCCATGCCAAAATCCATCCAAAATCCATC 107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TGTTTGGGAAATCTTCGTTAGTCCATCTCCATGCCAAAATCCATCCGGGAAATCCATC 60

Qy        108 ATGGACTGTGGGCTGCTCAGTGGATGCAGAGTTCATGATTACAAAGAAGAACCCAGA 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATGGAC--TGGGCTGCTCAGTGGATGCAGAGTTCATGATTACAAAGAAGAACCCCAGA 118

Qy        168 AGAGAAGAGACCTACTACATAGAATTATGGGAGTGTTGGAGGCTCGTGTGGGCAGCAG 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 AGAGAAGAGACCTACTACATAGAATTATGGGAGTGTTGGAGGCTCGTGTGGGCAGCAG 178

Qy        228 CGTGAAAAGCACAAGAGCAGTATTCACAACTCCGTAAATGTATTATTTCGTACACGA 287
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 CGTGAAAAGCACAAGAGCAGTATTCACAACTCCGTAAATGTATTATTTCGTACACGA 238

Qy        288 CTTAACAAATAAGAAGTCCTCCCAAAACTTGGTCGTTGGTCATTGGAAGTCTCAACAG 347
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        239 CTTAACAAATAAGAAGTCCTCCCAAAACTTGGTCGTTGGTCATTGGAAGTCTCAACAG 298

Qy        348 GGATTTGGTGCCAACTGGAGTCTTGGGTGACAAATGGGATTATGATCAAGAACAGTTGC 407
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 GGATTTGGTGCCAACTGGAGTCTTGGGTGACAAATGGGATTATGATCAAGAACAGTTGC 358

Qy        408 TGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCCAGATTCATGAAACAAA 467
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 TGATAACCAAATACCACTGTTGGTAATAGGGACTAAACTGGACCCAGATTCATGAAACAAA 418

Qy        468 GCGCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAAT 527
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 GCGCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAAT 478

Qy        528 TAATTTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCT 587
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 TAATTTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAAGCT 538

Qy        588 CAGTAGGTTTTTGATAAGGTCATAGAGAAGAGATAC-TTTTTAAGAGAAGAGGTAATCAGA 646
              ||||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db        539 CAGTAGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTTTAAGAGAAGAGGTAATCAGA 598

Qy        647 TTCCAGGCTTCCTGATCGG-AAAAGATTTGGGCAGGAACATTAAAGAGCCTTCATTAT 705
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        599 TTCCAGGCTTCCTGATCGGAAAAAGATTTGGGCAGGAACATTAAAGAGCCTTCATTAT 658

Qy        706 GACTGA 711
              ||||||
Db        659 GACTGA 664
```

Search completed: June 30, 2005, 16:53:34
Job time : 2655.06 secs

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      June 30, 2005, 14:24:03 ; Search time 3257.89 Seconds
                                       (without alignments)
                                       10574.845 Million cell updates/sec

Title:       US-09-945-173-3
Perfect score: 711
Sequence:    1 atggcgtccctggatcggt.........agagcctcattatgactga  711

Scoring table: IDENTITY_NUC
             Gapop 10.0 , Gapext 1.0
Searched:    4708233 seqs, 2422607955 residues

Total number of hits satisfying chosen parameters:        9415466

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
         1:  gb_ba:*
         2:  gb_htg:*
         3:  gb_in:*
         4:  gb_om:*
         5:  gb_ov:*
         6:  gb_pat:*
         7:  gb_ph:*
         8:  gb_pl:*
         9:  gb_pr:*
        10:  gb_ro:*
        11:  gb_sts:*
        12:  gb_sy:*
        13:  gb_un:*
        14:  gb_vi:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the total score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	711	100.0	711	6	AX465671	AX465671 Sequence
2	711	100.0	1694	6	AX465669	AX465669 Sequence
3	711	100.0	3362	9	HSM807705	BX647559 Homo sapi
4	711	100.0	3449	9	HSM807739	BX647593 Homo sapi
5	709.4	99.8	3352	9	HSM807810	BX647664 Homo sapi
6	707.8	99.5	1999	9	BC020832	BC020832 Homo sapi
7	587.8	82.7	2064	10	BC050194	BC050194 Mus muscu
8	484.2	68.1	3327	9	AJ720020	AJ720020 Gallus ga
9	435.4	61.2	2550	9	AK025772	AK025772 Homo sapi
10	418.2	58.8	1064	5	BC073035	BC073035 Xenopus l
11	396.6	55.8	1040	5	BC078191	BC078191 Danio rer
12	395	55.6	1291	5	AY391440	AY391440 Danio rer
13	365.8	51.4	385	6	BD076273	BD076273 5' EST of
14	223	31.4	223	6	CQ719549	CQ719549 Sequence
15	154.2	21.7	65838	9	AC133474	AC133474 Homo sapi
16	154.2	21.7	146010	9	AC009289	AC009289 Homo sapi
17	139.2	19.6	66009	2	AC101250	AC101250 Mus muscu
18	139.2	19.6	155418	2	AC117754	AC117754 Mus muscu
c 19	135	19.0	119730	9	AC117472	AC117472 Homo sapi
20	135	19.0	188485	2	AC027454	AC027454 Homo sapi
21	128.6	18.1	245326	2	AC107288	AC107288 Rattus no
22	128.6	18.1	266597	2	AC119722	AC119722 Rattus no
23	102.2	14.4	165836	2	AC138153	AC138153 Danio rer
24	102.2	14.4	174601	5	BX571724	BX571724 Zebrafish
25	102.2	14.4	180928	2	AC139727	AC139727 Danio rer
c 26	81.6	11.5	87017	2	AC140043	AC140043 Mus muscu
c 27	77.6	10.9	29949	2	AC112488	AC112488 Homo sapi
c 28	77.6	10.9	179351	2	AC143587	AC143587 Macaca mu
c 29	66.4	9.3	919	6	CQ581943	CQ581943 Sequence
30	66.4	9.3	934	6	AY070945	AY070945 Drosophil
31	66.4	9.3	1087	6	AR508948	AR508948 Drosophil
32	56.8	8.0	3048	6	CQ581942	CQ581942 Sequence
33	56.8	8.0	3215	6	CQ591803	CQ591803 Sequence
34	56.8	8.0	3888	6	CQ591806	CQ591806 Sequence
35	56.8	8.0	8610	3	AC012956	AC012956 Drosophil
c 36	56.8	8.0	177576	3	AC022349	AC022349 Drosophil
c 37	56.8	8.0	179077	3	AC018489	AC018489 Drosophil
38	56.8	8.0	300469	3	AE003503	AE003503 Drosophil
c 39	49.2	6.9	188485	2	AC027454	AC027454 Homo sapi
40	47.6	6.7	138	6	CQ736278	CQ736278 Sequence
41	46	6.5	1541	8	AK068217	AK068217 Oryza sat
42	45.2	6.4	1438	8	AK071529	AK071529 Oryza sat
c 43	45	6.3	66009	2	AC101250	AC101250 Mus muscu
c 44	44.4	6.2	183353	9	AC087071	AC087071 Homo sapi
45	44.4	6.2	185273	9	AC073320	AC073320 Homo sapi

ALIGNMENTS

```
RESULT 1
AX465671
LOCUS       AX465671                   711 bp    DNA        linear    PAT 16-JUL-2002
DEFINITION  Sequence 3 from Patent WO0218425.
ACCESSION   AX465671
VERSION     AX465671.1  GI:21899931
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Meyers,R.A.
  TITLE     47324, a human g-protein and uses therefor
  JOURNAL   Patent: WO 0218425-A 3 07-MAR-2002;
            Millennium Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1..711
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match             100.0%;  Score 711;  DB 6;  Length 711;
Best Local Similarity   100.0%;  Pred. No. 4.4e-198;
Matches  711;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1 ATGGCGTCCCTGGATCGGTGAAGGTACTGGTCGTTGGGAGACTTCAGGTGTTGGGAAATCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGCGTCCCTGGATCGGTGAAGGTACTGGTCGTTGGGAGACTTCAGGTGTTGGGAAATCT 60

Qy     61 TCGTTAGTCCATCTCCTATGCCAAAATCAAGTCGTCGGAAATCCATCATGGACTGTGGGC 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCGTTAGTCCATCTCCTATGCCAAAATCAAGTCGTCGGAAATCCATCATGGACTGTGGGC 120

Qy    121 TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAAGAACCCCAGAAGAGAGACCTAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAAGAACCCCAGAAGAGAGACCTAC 180

Qy    181 TACATAGAATTATGGGATGTTGGAGCTCGTCGTGGGCAGTGCCAGCAGCGTGAAAAGCACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TACATAGAATTATGGGATGTTGGAGCTCGTCGTGGGCAGTGCCAGCAGCGTGAAAAGCACA 240
```

Qy   19 ATGGGCTCCCTGGATCGGGTGAAGGTACTGTGTCGTGTTGGGAAATCT   78

Db   19 ATGGGCTCCCTGGATCGGGTGAAGGTACTGTGTCGTGTTGGGAAATCT 120

Qy   61 TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCATCATGGACTGTGGGC 120

Db   79 TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCATCATGGACTGTGGGC 138

Qy 121 TGCTCAGTGGATGTCAGAGTGTCATGATTACAAAGAAGGAACCCCAGAGAGAGACCTAC 180

Db 139 TGCTCAGTGGATGTCAGAGTGTCATGATTACAAAGAAGGAACCCCAGAGAGAGAAGACCTAC 198

Qy 181 TACATAGAATTATGGACAAATGGGATTGGAGGCTCTGTGGGCAGTGCCAGCGCGTGAAAAGCACA 240

Db 199 TACATAGAATTATGGACAAATGGGATTGGAGGCTCTGTGGGCAGTGCCAGCGCGTGAAAAGCACA 258

Qy 241 AGAGCAGTATTCTACACACTCCGTAAATGGTATTATTTTCGTACACGACTTAACACAAATAAG 300

Db 259 AGAGCAGTATTCTACACACTCCGTAAATGGTATTATTTTCGTACACGACTTAACACAAATAAG 318

Qy 301 AAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCA 360

Db 319 AAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCA 378

Qy 361 ACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATA 420

Db 379 ACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATA 438

Qy 421 CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGGCCATGAAGTT 480

Db 439 CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGGCCATGAAGTT 498

Qy 481 TTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGC 540

Db 499 TTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGC 558

Qy 541 ACAAATCCACGGTACTTAGCTCGACTGCAGGTTCTTCCAATGCTGTCTCAAGCTCAGTAGGTTTTT 600

Db 559 ACAAATCCACGGTACTTAGCTCGACTGCAGGTTCTTCCAATGCTGTCTCAAGCTCAGTAGGTTTTT 618

Qy 601 GATAAGGTCATAGAGAGAAGAGATACTTTTAAGAGAGAAGGTAATCAGATTCAGATTCCAGGCTTTCCT 660

Db 619 GATAAGGTCATAGAGAGAAGAGATACTTTTAAGAGAGAAGGTAATCAGATTCAGATTCCAGGCTTTCCT 678

Qy 661 GATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGCCTTCATTATGACTGA 711

Db 679 GATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGCCTTCATTATGACTGA 729

RESULT 3
HSM807705
LOCUS   HSM807705   3362 bp   mRNA   linear   PRI 30-AUG-2003
DEFINITION Homo sapiens mRNA; cDNA DKFZp313A1611 (from clone DKFZp313A1611).
ACCESSION   BK647559
VERSION   BK647559.1   GI:34366716
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 3362)
  AUTHORS   Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  CONSRTM   The German Human cDNA Consortium
  TITLE   Direct Submission
  JOURNAL   Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT   Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by BMFZ (Biomedical Research Center at the
            Heinrich-Heine-University, Duesseldorf/Germany) within the cDNA
            sequencing consortium of the German Genome Project. This clone
            (DKFZp313A1611) is available at the RZPD in Berlin. Please contact
            the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further

RESULT 2
AX465669
LOCUS   AX465669   1694 bp   DNA   linear   PAT 16-JUL-2002
DEFINITION Sequence 1 from Patent WO0218425.
ACCESSION   AX465669
VERSION   AX465669.1   GI:21899929
KEYWORDS   .
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Meyers,R.A.
  TITLE   47324, a human g-protein and uses therefor
  JOURNAL   Patent: WO 0218425-A 1 07-MAR-2002;
            Millennium Pharmaceuticals (US)
FEATURES   Location/Qualifiers
     source   1..1694
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     CDS   19..729
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD42379.1"
                     /db_xref="GI:21899930"
                     /translation="MASLDRVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVD
                     VRVHDYKEGTPEKKTYYIEBLMDVGGSVGSASSVKSTRAVFYNSVNGIIFVHDLTNKKS
                     SQNLRRWSLEALNRDLVPTGVLVTNGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEV
                     LRTAFLAEDFNPEEINLDCTNPRYLAAGSSNAVKLSRFFDKVIEKRYFLREGNQIPG
                     FPDRKRFGAGTLKSLHYD"
ORIGIN
Query Match   100.0%; Score 711; DB 6; Length 1694;
Best Local Similarity   100.0%; Pred. No. 4.5e-198;
Matches 711; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 ATGGGCTCCCTGGATCGGGTGAAGGTACTGTGTCGTGTTGGGAAATCT 60

Qy 241 AGAGCAGTATTCTACACACTCCGTAAATGGTATTATTTTCGTACACGACTTAACACAAATAAG 300

Db 241 AGAGCAGTATTCTACACACTCCGTAAATGGTATTATTTTCGTACACGACTTAACACAAATAAG 300

Qy 301 AAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCA 360

Db 301 AAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCA 360

Qy 361 ACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATA 420

Db 361 ACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATA 420

Qy 421 CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGGCCATGAAGTT 480

Db 421 CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGGCCATGAAGTT 480

Qy 481 TTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGC 540

Db 481 TTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGC 540

Qy 541 ACAAATCCACGGTACTTAGCTCGACTGCAGGTTCTTCCAATGCTGTCTCAAGCTCAGTAGGTTTTT 600

Db 541 ACAAATCCACGGTACTTAGCTCGACTGCAGGTTCTTCCAATGCTGTCTCAAGCTCAGTAGGTTTTT 600

Qy 601 GATAAGGTCATAGAGAGAAGAGATACTTTTAAGAGAGAAGGTAATCAGATTCAGATTCCAGGCTTTCCT 660

Db 601 GATAAGGTCATAGAGAGAAGAGATACTTTTAAGAGAGAAGGTAATCAGATTCAGATTCCAGGCTTTCCT 660

Qy 661 GATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGCCTTCATTATGACTGA 711

Db 661 GATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGCCTTCATTATGACTGA 711

```
ACCESSION   BX647593
VERSION     BX647593.1  GI:34366750
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Theria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3449)
  AUTHORS   Koehrer,K., Beyer,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A.,
            Fobo,G., Han,M. and Wiemann,S.
  CONSRTM   The German Human cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by BMFZ (Biomedical Research Center at the
            Heinrich-Heine-University, Duesseldorf/Germany) within the cDNA
            sequencing consortium of the German Genome Project. This clone
            (DKFZp313M1333) is available at the RZPD in Berlin. Please contact
            the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY, Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES             Location/Qualifiers
     source          1..3449
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp313M1333"
                     /tissue_type="human heart"
                     /clone_lib="313 (synonym: hlcc2). Vector pTriplEx2; host
                     DH10B; sites SfiIA + SfiIB"
                     /dev_stage="adult"
     polyA_signal    3311..3316
     polyA_site      3336
ORIGIN

Query Match      100.0%;      Score 711;  DB 9;  Length 3449;
Best Local Similarity 100.0%;   Pred. No. 4.7e-198;
Matches   711;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy    1 ATGGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCT 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   31 ATGGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCT 90

Qy   61 TCGTTAGTCCATCTCCATGCCAAAATCAAGTGCTGGGAAATCCATCATGCGACTGTGGGC 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   91 TCGTTAGTCCATCTCCATGCCAAAATCAAGTGCTGGGAAATCCATCATGCGACTGTGGGC 150

Qy  121 TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAGAAGACCTAC 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  151 TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAGAAGACCTAC 210

Qy  181 TACATAGAATTATGGGATGTTGGAGGCTCTCGTGGGCAGTGCCAGCAGCGTGAAAAGCACA 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  211 TACATAGAATTATGGGATGTTGGAGGCTCTCGTGGGCAGTGCCAGCAGCGTGAAAAGCACA 270

Qy  241 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTAACAAATAAG 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  271 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTAACAAATAAG 330

Qy  301 AAGTCCTCCCAAAACTTGCGTCGTTCGGTCGTTGGTCATTGGAAGCTCTCAACAGGGATT 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  331 AAGTCCTCCCAAAACTTGCGTCGTTCGGTCGTTGGTCATTGGAAGCTCTCAACAGGGATT 390

Qy  361 ACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAAACAGTTTGCTGATAACCAAATA 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  391 ACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAAACAGTTTGCTGATAACCAAATA 450

Qy  421 CCACTGTTGGTAATAGGGACTAAACTGACCAGAGTTCATGGAAAACAAAGCGCCATGAAGTT 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  451 CCACTGTTGGTAATAGGGACTAAACTGACCAGAGTTCATGGAAAACAAAGCGCCATGAAGTT 510
```

FEATURES   Location/Qualifiers
     source   1..3362
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp313A1611"
                     /tissue_type="human bone marrow"
                     /clone_lib="313 (synonym: hlcc2). Vector pTriplEx2; host
                     DH10B; sites SfiIA + SfiIB"
                     /dev_stage="adult"
     polyA_signal   3308..3313
     polyA_site   3332
ORIGIN

Query Match   100.0%;   Score 711; DB 9; Length 3362;
Best Local Similarity 100.0%;   Pred. No. 4.7e-198;
Matches   711;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;
```

```
Qy    1 ATGGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCT 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   28 ATGGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCT 87

Qy   61 TCGTTAGTCCATCTCCATGCCAAAATCAAGTGCTGGGAAATCCATCATGCGACTGTGGGC 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   88 TCGTTAGTCCATCTCCATGCCAAAATCAAGTGCTGGGAAATCCATCATGCGACTGTGGGC 147

Qy  121 TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAGAAGACCTAC 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  148 TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAGAAGACCTAC 207

Qy  181 TACATAGAATTATGGGATGTTGGAGGCTCTCGTGGGCAGTGCCAGCAGCGTGAAAAGCACA 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  208 TACATAGAATTATGGGATGTTGGAGGCTCTCGTGGGCAGTGCCAGCAGCGTGAAAAGCACA 267

Qy  241 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTAACAAATAAG 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  268 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTAACAAATAAG 327

Qy  301 AAGTCCTCCCAAAACTTGCGTCGTTCGGTCGTTGGTCATTGGAAGCTCTCAACAGGGATT 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  328 AAGTCCTCCCAAAACTTGCGTCGTTCGGTCGTTGGTCATTGGAAGCTCTCAACAGGGATT 387

Qy  361 ACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAAACAGTTTGCTGATAACCAAATA 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  388 ACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAAACAGTTTGCTGATAACCAAATA 447

Qy  421 CCACTGTTGGTAATAGGGACTAAACTGACCAGAGTTCATGGAAAACAAAGCGCCATGAAGTT 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  448 CCACTGTTGGTAATAGGGACTAAACTGACCAGAGTTCATGGAAAACAAAGCGCCATGAAGTT 507

Qy  481 TTAACTAGGACTGCTTCCTGGCTGAGGATTCAATCAGAGAAGAAATTAATTGGACTGC 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 TTAACTAGGACTGCTTCCTGGCTGAGGATTCAATCAGAGAAGAAATTAATTGGACTGC 540

Qy  508 TTAACTAGGACTGCTTCCTGGCTGAGGATTCAATCCAGAGAAGAAATTAATTGGACTGC 567

Qy  541 ACAAATCCACGTACTTAGCTGCAGGTTCTCCAAATGCTGTCAAGCTCAGTAGGTTTTTT 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  568 ACAAATCCACGTACTTAGCTGCAGGTTCTCCAAATGCTGTCAAGCTCAGTAGGTTTTTT 627

Qy  601 GATAAGGTCATAGAAGAAGATACTTTTAAGAGAAGGTAATCAGATTCAGATTCAGATTATGACTGA 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  628 GATAAGGTCATAGAAGAAGATACTTTTAAGAGAAGGTAATCAGATTCAGATTCAGATTATGACTGA 687

Qy  661 GATCGGAAAAGATTTGGGCGCAGGAACATTAAAGAGCCTTCATTATGACTGA 711
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  688 GATCGGAAAAGATTTGGGCGCAGGAACATTAAAGAGCCTTCATTATGACTGA 738
```

```
RESULT 4
HSM807739
LOCUS       HSM807739   3449 bp   mRNA   linear   PRI 30-AUG-2003
DEFINITION  Homo sapiens mRNA; cDNA DKFZp313M1333 (from clone DKFZp313M1333).
```

```
Qy  121 TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAGGAACCCCAGAAGAGAAGACCTAC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  135 TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAGGAACCCCAGAAGAGAAGACCTAC 194

Qy  181 TACATAGAATTATGGACGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGAAAAGCACA 240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  195 TACATAGAATTATGGACGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGAAAAGCACA 254

Qy  241 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGACTTAACAAATAAG 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  255 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGACTTAACAAATAAG 314

Qy  301 AAGTCCTCCCAAAACTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGTGCCA 360
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  315 AAGTCCTCCCAAAACTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGTGCCA 374

Qy  361 ACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAAGAGTTGCTGATAACCAAATA 420
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  375 ACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAAGAGTTGCTGATAACCAAATA 434

Qy  421 CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTT 480
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  435 CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTT 494

Qy  481 TTAACTAGGACTGCTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGC 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  495 TTAACTAGGACTGCTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGC 554

Qy  541 ACAAATCCACGGTACTTAGCTGCAGTTCTTCCAATGCTGTCAGCTCAGTAGGTTTTTT 600
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  555 ACAAATCCACGGTACTTAGCTGCAGTTCTTCCAATGCTGTCAGCTCAGTAGGTTTTTT 614

Qy  601 GATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAATCAGATTCCAGGCTTTCCT 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  615 GATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAATCAGATTCCAGGCTTTCCT 674

Qy  661 GATCGGAAAAGATTTGGGCAGGAGGAACATTAAAGAGCCTTCATTATGACTGA 711
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  675 GATCGGAAAAGATTTGGGCAGGAGGAACATTAAAGAGCCTTCATTATGACTGA 725
```

```
RESULT 6
BC020832
LOCUS       BC020832               1999 bp    mRNA    linear   PRI 29-JUN-2004
DEFINITION  Homo sapiens RAB, member of RAS oncogene family-like 3, mRNA (cDNA
            clone MGC:23920 IMAGE:4771781), complete cds.
ACCESSION   BC020832
VERSION     BC020832.1  GI:18088571
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1999)
AUTHORS     Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE       Generation and initial analysis of more than 15,000 full-length
```

```
Qy  481 TTAACTAGGACTGCTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGC 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  511 TTAACTAGGACTGCTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGC 570

Qy  541 ACAAATCCACGGTACTTAGCTGCAGTTCTTCCAATGCTGTCAGCTCAGTAGGTTTTTT 600
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  571 ACAAATCCACGGTACTTAGCTGCAGTTCTTCCAATGCTGTCAGCTCAGTAGGTTTTTT 630

Qy  601 GATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAATCAGATTCCAGGCTTTCCT 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  631 GATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAATCAGATTCCAGGCTTTCCT 690

Qy  661 GATCGGAAAAGATTTGGGCAGGAGGAACATTAAAGAGCCTTCATTATGACTGA 711
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  691 GATCGGAAAAGATTTGGGCAGGAGGAACATTAAAGAGCCTTCATTATGACTGA 741
```

```
RESULT 5
HSM807810
LOCUS       HSM807810               3352 bp    mRNA    linear   PRI 30-AUG-2003
DEFINITION  Homo sapiens mRNA; cDNA DKFZp3130171 (from clone DKFZp3130171).
ACCESSION   BX647664
VERSION     BX647664.1  GI:34366821
KEYWORDS    
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3352)
AUTHORS     Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
CONSRTM     The German Human cDNA Consortium
TITLE       Direct Submission
JOURNAL     Submitted (27-AUG-2003) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de;
            sequenced by DKFZ (German Cancer Research Center,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFZp3130171) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://mips.gsf.de/proj/cDNA/.
FEATURES             Location/Qualifiers
     source          1..3352
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp3130171"
                     /tissue_type="human fetal skin"
                     /clone_lib="313 (synonym: hlcc2). Vector pTripl8x2; host
                     DH10B; Sites SfiIA + SfiIB"
                     /dev_stage="fetal"
     polyA_signal    3296..3301
     polyA_site      3321
ORIGIN

Query Match      99.8%;  Score 709.4;  DB 9;  Length 3352;
Best Local Similarity  99.9%;  Pred. No. 1.4e-197;
Matches  710;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy    1 ATGGGCGTCCCTGGATCGGGTGAAGGTGACTGGTGTGTGTTGGGAGACTCAGGTGTTGGGAAATCT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   15 ATGGCGTCCCTGGATCGGGTGAAGGTGACTGGTGTGTGTTGGGAGACTCAGGTGTTGGGAAATCT 74

Qy   61 TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGGC 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75 TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGGC 134
```

```
Db  190 TACATAGAATTATGGATGTTGGAGGCTCTGTGTCGGGCAGTGCCAGCAGCGTGAAAAGCACA 249
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  241 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTAACAAATAAG 300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  250 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTAACAAATAAG 309
Qy  301 AAGTCCTCCCAAACTTGGTCGTCGTTGGGTCATTCGAAGCTTCAACAGGGGATTTGGTGTGCCA 360
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  310 AAGTCCTCCCAAACTTGCGTCGTCGTTGGGTCATTCGAAGCTTCAACAGGGATTTGGTGTGCCA 369
Qy  361 ACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGAACAGTTGCTGATAACCAAATA 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  370 ACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGAACAGTTGCTGATAACCAAATA 429
Qy  421 CCACTGTTGGTATAGGGACTAAACTGACCAGATTCATGAAACAAAGCGCCATGAAGTT 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  430 CCACTGTTGGTATAGGGACTAAACTGACCAGATTCATGAAACAAAGCGCCATGAAGTT 489
Qy  481 TTAACTAGGACTGCTTTCCTGCGTCGAGATTTCAATCCAGAAGAAAATTAATTTGGACTGC 540
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  490 TTAACTACGACTGCTTTCCTGCGTCGAGATTTCAATCCAGAAGAAAATTAATTTGGACTGC 549
Qy  541 ACAAATCCACGTACTAGCTGCAGTTCTTCCAGTGTCGAAGCTCGAAGCTCAGTAGGTTTTT 600
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  550 ACAAATCCACGTACTTAGCTGCAGTTCTTCCAGTGTCGAAGCTCGAAGCTCAGTAGGTTTTT 609
Qy  601 GATAAGGTCATAGAGAAGAGATACTTTTAAGAGAAGAAGGTAATCAGATTCCAGGCTTTCCT 660
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  610 GATAAGGTCATAGAGAAGAGATACTTTTAAGAGAAGAAGGTAATCAGATTCCAGGCTTTCCT 669
Qy  661 GATCGGAAAAGATTTGGGCAGGCAGGAACATTAAAGAGCCTTCATTATGACTGA 711
        |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  670 GATCGGAAAAGATTTGGGCAGGCAGGAACACATTAAAGAGCCTTCATTATGACTGA 720
```

```
RESULT 7
BC050194
LOCUS       BC050194                2064 bp    mRNA    linear    ROD 08-OCT-2003
DEFINITION  Mus musculus RAB, member of RAS oncogene family-like 3, mRNA (cDNA
            clone MGC:60441 IMAGE:30022531), complete cds.
ACCESSION   BC050194
VERSION     BC050194.1  GI:29747934
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2064)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
  PUBMED    12477932
REFERENCE   2  (bases 1 to 2064)
```

```
  JOURNAL   human and mouse cDNA sequences
            Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  PUBMED    12477932
REFERENCE   2  (bases 1 to 1999)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JAN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
  COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA 94305
            Web site:     http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson,M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
            R. M.
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 37 Row: m Column: 23
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 28376659.
FEATURES             Location/Qualifiers
     source          1..1999
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:23920 IMAGE:4771781"
                     /tissue_type="Testis, embryonal carcinoma"
                     /clone_lib="NIH_MGC_61"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
     gene            1..1999
                     /gene="RABL3"
                     /note="synonym: MGC23920"
                     /db_xref="LocusID:285282"
     CDS             10..720
                     /gene="RABL3"
                     /codon_start=1
                     /product="RAB, member of RAS oncogene family-like 3"
                     /protein_id="AAH20832.1"
                     /db_xref="GI:18088572"
                     /db_xref="LocusID:285282"
                     /translation="MASLDRVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVD
                     VRVHDYKEGTPERKTCYIELWDVGGSVGGSASSVKSTRAVFYNSVNGIIFVHDLTNKKS
                     SQNLRRWSLEALNRDLVPTGVLVTNGDYDQEQFADNQIPLLVIGTKLDQIHETKRHEV
                     LTTTAFLASDFNPEINLDCTNPRYLAAGSSNAVKLSRFPDKVIEKRYFLREGNQIPG
                     FPDRKRFGAGTLKSLHYD"
ORIGIN
Query Match    99.5%;  Score 707.8;  DB 9;  Length 1999;
Best Local Similarity  99.7%;  Pred. No. 4e-197;
Matches  709;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy    1 ATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10 ATGGCGGCCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCT 69
Qy   61 TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGACTGTGGGC 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70 TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGACTGTGGGC 129
Qy  121 TGCTAGTCAGTGGATGTCAGAGTTCATGATTACAAGAAGGAACCCCAGAAGAAGACCTAC 180
        |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  130 TGCTCAGTGGATGTCAGAGTTCATGATTACAAGAAGGAACCCCAGAAGAAGACCTGC 189
Qy  181 TACATAGAATTATGGGATGTTGGAGGCTCTGTGTCGGGCAGTGCCAGCAGCGTGAAAAGCACA 240
```

```
Db   63  TCACTCGTCCATCTTCTGTGCCACAATCAGTGTTAGGAAATCCGTCATGGACTGTGGGC 122
             ||||| |||||||| || |||||||||||||||| || |||||||||||||||||||
Qy  121  TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAGAGGAACCCCAGAAGAGAGACTTAC 180
             |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  123  TGCTCCGGTGGATATCAGAGTTCATGATTACAAAGAGAGGAACCCCCTGAAGAAGACATAC 182

Qy  181  TACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCGCGTGAAAAGCACA 240
             || |||||| ||||| ||| ||||||||||| |||||||||||||||||||||||||||
Db  183  TATATAGAACTATGGGATGTTGGAGGCTCTGTGGGCCAGTGCCAGCAGTGTGAAAAGCACA 242

Qy  241  AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGACTTAACAAATAAG 300
             ||||||||||||||||||| ||| |||| |||||||||| |||| ||||||||||||||
Db  243  AGAGCGGTGTTCTACAACTCTGTAAACGGCATCATTTAGTACATGACTTAACAAATAAG 302

Qy  301  AAGTCCTCCCAAAACTGCCTCGTTGGTCCGGTGGAAGCTCTCAACAGGGATTTGGTGCCA 360
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  303  AAGTCATCTCAAAACTTATATCGTCGGTGGTCATTGGTTCATTGGAAGTTCCAATCGGGAT 362

Qy  361  ACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATA 420
             ||||||||||||||||| |||| |||||||||||||||||||| ||||||||||||||||
Db  363  ACCGGAGTCCTGGTGACAAATGGGGATTATGACCGAGAACAGTTTGCTGATAACCAAATC 422

Qy  421  CCACTGTTGTAATAGGGACTAAACTGGACCAGATTCATGAAAACAAGCGCCATGAAGTT 480
             |||||||||||||||| |||||||||||||||||| ||||||||||||||||||||||||
Db  423  CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAAACCAAGCGCCATGAAGTT 482

Qy  481  TTAACTAGGACTGCTTCCTGGCTGAGGAGATTCAATCCAGAAGAAATTAATTTGGACTGC 540
             |||| |||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db  483  TTAATCAGGACTGCTTTCTGGCTGAGGAGACTTCAATGCTGAAGAGATTAATTTGGATTGC 542

Qy  541  ACAAATCCACGGTACTTAGCTGCAGTTCTTCCAAATGCTGTCAAGCTCAGTAGGTTTTTT 600
             |||| ||| ||| |||||| ||| || ||||||  ||||| |||  || |||||||||||
Db  543  ACAAACCCACCGGTCCTCAGCTGCAGGCTCCTCCAATGCTGTCAAGCTCGTCAGTAGATTTTTTT 602

Qy  601  GATAAGGTCATAGAGAAGAGATACTTTTAAGAGAAGGTAATCAGATTCCAGGCTTCCT 660
             |||||||||||||||||||||||||||||||| |||| ||||| |||||||||||||| 
Db  603  GATAAGGTTATAGAAGAGAGATATTTTTTTAGAGAAGGTAATCAGATTCCGGGCTTTTCT 662

Qy  661  GATCGGGAAAAGATTTGGGGCAGGAGGACCTTCATTATGCACTGA 711
             ||||||||||||| || ||| ||||  |||||||||||||||||||
Db  663  GATCGGGAAGAGGTTTGGGAGGAGGGCACTCGAAGAACTTCCACTGTGACTGA 713
```

```
RESULT 8
AJ720020
LOCUS       AJ720020                3327 bp    mRNA    linear   VRT 30-SEP-2004
DEFINITION  Gallus gallus mRNA for hypothetical protein, clone 9j2.
ACCESSION   AJ720020
VERSION     AJ720020.1  GI:53130699
KEYWORDS    ORF1.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1
  AUTHORS   Caldwell,R.B., Kierzek,A.M., Arakawa,H., Bezzubov,Y., Zaim,J.,
            Fiedler,P., Kutter,S., Blagodatski,A., Kostovska,D., Koter,M.,
            Plachy,J., Carninci,P., Hayashizaki,Y. and Buerstedde,J.M.
  TITLE     Full-length cDNAs from bursal lymphocytes to facilitate gene
            function analysis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3327)
  AUTHORS   Caldwell,R.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-2004) Caldwell R.B., GSP - Forschungszentrum,
            Institut fuer Molekulare Strahlenbiologie, Ingolstaedter Landstr.
            1, D-85764 Neuherberg, GERMANY
FEATURES             Location/Qualifiers
     source          1..3327
                     /organism="Gallus gallus"
                     /mol_type="mRNA"
                     /strain="CB"
```

AUTHORS   Strausberg,R.
TITLE   Direct Submission
JOURNAL   Submitted (02-APR-2003) National Institutes of Health, Mammalian
           Gene Collection (MGC), Cancer Genomics Office, National Cancer
           Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
           USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
           Email: cgapbs-r@mail.nih.gov
           Tissue Procurement: Dr. Janet Rossant and Tilo Kunath (Samuel
           Lunenfeld Research Institute, Canada)
           cDNA Library Preparation: Yulan Piao and Minoru Ko (National
           Institute on Aging, NIH: http://lgsun.grc.nia.nih.gov/cDNA/)
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Baylor College of Medicine Human Genome
           Sequencing Center
           Center code: BCM-HGSC
           Web site: http://www.hgsc.bcm.tmc.edu/cdna/
           Contact: amg@bcm.tmc.edu
           Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
           Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
           A.N., Gibbs, R.A.
           Clone distribution: MGC clone distribution information can be found
           through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
           Series: IRAK Plate: 111 Row: m Column: 3
           This clone was selected for full length sequencing because it
           passed the following selection criteria: matched mRNA gi: 13385791.

```
FEATURES             Location/Qualifiers
     source          1..2064
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="B5/EGFP transgenic ICR mice"
                     /db_xref="taxon:10090"
                     /clone="MGC:60441 IMAGE:30022531"
                     /tissue_type="Trophoblast Stem Cell, mouse, 3.5 dpc"
                     /clone_lib="NIA Mouse Trophoblast Stem Cell cDNA Library
                     (Long)"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1"
     gene            1..2064
                     /gene="Rab13"
                     /db_xref="LocusID:67657"
                     /db_xref="MGI:1914907"
     CDS             3..713
                     /gene="Rab13"
                     /codon_start=1
                     /product="Rab13 protein"
                     /protein_id="AAH50194.1"
                     /db_xref="GI:29747935"
                     /db_xref="LocusID:67657"
                     /db_xref="MGI:1914907"
                     /translation="MASLDRVKVLVLGDSGVGKSSLVHLLCHNQVLGNPSWTVGCSVD
                     IRVHDYKRGTPEEKTYIELWDVGGSVGSASSVKSTRAVFYNSVNGIILVHDLTNKKS
                     SQNLYRWSLEVLNRDAVPTGVLTVNGDYDREQFADNQIPLLVIGTKLDQIHETKRHEV
                     LIRTAFLABEDFNAEEINLDCTNPRSSAAGSSNAVKLSRFPFDKVIEKRYFFREGNQIPG
                     FSDRKRFGGGALKNFHCD"
     misc_feature    21..464
                     /gene="Rab13"
                     /note="RAB; Region: Rab subfamily of small GTPases"
                     /db_xref="CDD:smart00175"
ORIGIN
```

```
Query Match        82.7%;  Score 587.8;  DB 10;  Length 2064;
Best Local Similarity  89.2%;  Pred. No. 9e-162;
Matches  634;  Conservative  0;  Mismatches  77;  Indels  0;  Gaps  0;

Qy    1  ATGGCGTCCCTGGATCGGGTGAAGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAAATCT 60
            |||||||||| |||||||||||||||||||||||||| |||||||||||||||||||||
Db    3  ATGGCGTCCCTGGATCGGGTGAAGGTTCTCGTGTTGGGAGACTCAGGTGTTGGGAAAATCT 62

Qy   61  TCGTTAGTCATCCTCTATGCCAAAATCAAGTCAGTGGAAATCCATCATCGGACTGTGGGC 120
            || ||||||| |||| ||||||||||||||||| ||||||| |||||||||||||||||
Db   61  ...
```

```
Qy    659  CTGATCGGAAAAGAGATTTGGGCGCAGGAGGAACATTAAAGAGCCTTCATTATGACTGA   711
                      |||||||||||||||| ||||||||||| ||||||||||||||||||||||
Db    691       CGGAAAGAAAGAGATTTGGAGGAGGAACACTGAAGAGCCTTCATTATGACTGA   743


RESULT 9
AK025772
LOCUS       AK025772                2550 bp    mRNA    linear   PRI 13-SEP-2003
DEFINITION  Homo sapiens cDNA: FLJ22119 fis, clone HEP18852.
ACCESSION   AK025772
VERSION     AK025772.1  GI:10438390
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y.,
            Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
            Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2550)
  AUTHORS   Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
            Shibahara,T., Tanaka,T. and Nakamura,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
            University of Tokyo, Laboratory of Genome Structure Analysis, Human
            Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
            Japan (E-mail:flcdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
            Fax:81-3-5449-5416)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            International Trade and Industry of Japan; cDNA full insert
            sequencing: Research Association for Biotechnology; cDNA library
            construction, 5'- & 3'-end one pass sequencing: Depart of
            Virology and Human Genome Center, Institute of Medical Science,
            University of Tokyo (partly supported by Science and Technology
            Agency).
FEATURES             Location/Qualifiers
     source          1..2550
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="HEP18852"
                     /cell_line="HepG2"
                     /cell_type="hepatoma"
                     /clone_lib="HEP"
                     /note="cloning vector pME18SFL3"
ORIGIN

Query Match             61.2%;  Score 435.4;  DB 9;  Length 2550;
Best Local Similarity   99.8%;  Pred. No. 7.2e-117;
Matches   436;  Conservative    0;  Mismatches   1;  Indels   0;  Gaps    0;

Qy    275  TTTTCGTACACGACTAACAAATAAGAAGTCCTCCAAAAACTTGCGTCGTTGGTCATTGG   334
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  TTTTCGTACACGACTAACAAATAAGAAGTCCTCCAAAAACTTGCGTCGTTGGTCATTGG    60

Qy    335  AAGCTCTCAACAGGGATTTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGATC   394
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  AAGCTCTCAACAGGGATTTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGATC   120

Qy    395  AAGAACAGTTTGCTGATAACCAAATACCACTGTTGTAATAGGGACTAAACTGGACCAGA   454
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  AAGAACAGTTTGCTGATAACCAAATACCACTGTTGTAATAGGGACTAAACTGGACCAGA   180

Qy    455  TTCATGAAACAAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCA   514
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  TTCATGAAACAAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCA   240

Qy    515  ATCCAGAAGAAATTAATTGGACTGCACAAATCCACGTGACTTAGCTGCAGGTCTTCCA   574
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
                     /db_xref="taxon:9031"
                     /clone="9j2"
                     /cell_type="lymphocyte"
                     /tissue_type="bursa"
                     /clone_lib="riken1"
                     /dev_stage="2 weeks old"
     gene            1..3327
                     /locus_tag="RCJMB04_9j2"
     5'UTR           <1..30
                     /locus_tag="RCJMB04_9j2"
     CDS             31..723
                     /locus_tag="RCJMB04_9j2"
                     /note="ORF1"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAG31679.1"
                     /db_xref="GI:53130700"
                     /translation="MAALDRVKVLVLGDSGVGKSSLVHLLCHNQVLGNPSWTVGCSVD
                     VRIHDYKGTPERKTYYIELWDVGGSVGSATSVKSTRAVPYNSLNGIILVHDLTNKKS
                     SQNLYRWSLEALNRDVAPTGVLVTNGDYDREQFADNQIPLLVIGTKLDQIPETKRNEV
                     LRTRTAFLABDFNABEINLDCTNPRYLAAGSSNAVKLSRFFDKVIEKRYPFKRWQSRPL
                     ASRKERDLEEBH"
     3'UTR           724..>3327
                     /locus_tag="RCJMB04_9j2"
ORIGIN

Query Match             68.1%;  Score 484.2;  DB 5;  Length 3327;
Best Local Similarity   81.8%;  Pred. No. 3.1e-131;
Matches   583;  Conservative    0;  Mismatches  128;  Indels   2;  Gaps    2;

Qy      1  ATGGGCGTCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAAATCT    60
           ||||| || |||||| ||||| |||||| |||||||||||| ||||| ||||| ||| |||
Db      1  ATGGGGGCTCTGGATAGGGTCAAGGTGCTGGTGTTGGGCGACTCCGGTGTCGGGAAAGTCT    90

Qy     61  TCGTTAGTCCATCCTATGCCAAAATCAAGTGCTCGGGAAATCAATCATCATGACTGTGGGC   120
           || |  ||| |||| ||| ||| | |||| ||| |||| | |||||  | ||||  ||||
Db     91  TCGCTCGTGCACCTCCTGTGTCACAACCAGGTGCTGGGGAACCCGTCCTGGACGGTGGGGC   150

Qy    121  TGCTCAGTGATGTCAGAGTTCATGATTACAAAGAAGAAGACCTAC                  180
           |||||  ||| ||| |||||||||| ||| |||  ||| ||| ||||
Db    151  TGCTCGGTGATGTGCGAATCCATGACTACAAAGGAGAAGAGAAGACATAC             210

Qy    181  TACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCGCACA            240
           || ||||| |||||||| || ||||| ||||  |||||| |||| |||||
Db    211  TATATAGAGCTGTGGGATGTTGGAGGTTCAGTGGGTAGTGCCACTAGTGTGAAAAGCACA   270

Qy    241  AGAGCAGTATTCTACAACTCCGTAAAATGGTATTATTTCGTACACGACTAACAAATAAG    300
           ||||||| || ||| |||| ||  ||||| | |  |||| ||
Db    271  AGAGCAGTATTTTATAACTCGCTGAATGGGATAATTTAGTGCATGCATGACTTAACCACAAG  330

Qy    301  AAGTCTCCCAAAACTTGCGTCGTGGTCGCTGGAGCTCTCAACAGGGATTTGGTGCCA      360
           ||   ||| |||| ||||| || || ||  |||||| || |||||| ||  ||||||
Db    331  AAATCATCCCAGAGAATTTATATCGCTGGTTCTTGGGAGCACTCAACAGAGATGTCGCTCCC  390

Qy    361  ACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATA   420
           |||||||||  |||||||  |||| || |||  |||| ||||| |||||||| ||||
Db    391  ACAGGAGTTCTGGTGACTAAACGGTATCAGGTCTGCAGCGGTCTTAACTGTCAAACCAGATT  450

Qy    421  CCACTGTTGTAATAGGGACTAAACAAAGCGCCATGAAGTTTAAC                  480
           |||||| | ||||  ||||| | |||| |||||||  ||| ||||
Db    451  CCACTGCTGTAATAGGGACTAAAGCGTAATGAAGTTTAAC                       510

Qy    481  TTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATTGGACTGC   540
           || ||| |||||| ||| ||||||||  | | ||   ||||  |||| |||||||||||
Db    511  TTGACCAGGACAGCTTTCTTGGCTGAGGATTTCAATGCTGAAGAGATAAATTTGGATTGC   570

Qy    541  ACAAATCCACGTGACTTAGCTGCAGGTCTTCCAATGCTGTCAAGCTCAGTAGTTTTTT   600
           || |||||| ||| || | |||||||||| ||  ||  ||  |||| ||||| |||||
Db    571  ACCAATCCTCTGCTGGCGCAGGCTCATCCATCAAACTAAGCAGCAGATTTTTTT        630

Qy    601  GATAAGGTCATAGAAGAGATACTT-TTTAAGAAGAAGGTAATC-AGATTCCAGGCTTTC   658
           ||| || || || ||||| |||||| || ||||||||||||   | |||||||| |||
Db    631  GATAAGGTGATAGAGAAGAGATACTTTCTTTAAGAGATGGCAATCAGATTCCTGGCTTCT   690
```

```
Db    241 ATCCAGAAGAAATTAATTTGGACTGCACAAATCCGCGGTACTTAGCTGCAGGTTCTTCCA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    575 ATGCTGTCAAGCTCAGTAGGTTTTTGATAAGGTCATCATGAGAGAAGAGATACTTTTAAGAG 634
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ATGCTGTCAAGCTCAGTAGGTTTTTGATAAGGTCATCATGAGAGAAGAGATACTTTTTAAGAG 360

Qy    635 AAGGTAATCAGATTCCAGGCTTCCTGATCGGAAAAGATTTGGGCAGGAACATTAAAGA 694
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AAGGTAATCAGATTCCAGGCTTCCTGATCGGAAAAGATTTGGGCAGGAACATTAAAGA 420

Qy    695 GCCTTCATTATGACTGA 711
           |||||||||||||||||
Db    421 GCCTTCATTATGACTGA 437
```

RESULT 10

```
BC073035
LOCUS       BC073035                1064 bp    mRNA    linear   VRT 03-AUG-2004
DEFINITION  Xenopus laevis MGC82648 protein, mRNA (cDNA clone MGC:82648
            IMAGE:5049322), complete cds.
ACCESSION   BC073035
VERSION     BC073035.1  GI:49257300
KEYWORDS    MGC.
SOURCE      Xenopus laevis (African clawed frog)
ORGANISM    Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus; Xenopus.
REFERENCE   1  (bases 1 to 1064)
AUTHORS     Klein,S.L., Strausberg,R.L., Wagner,L., Pontius,J., Clifton,S.W.
            and Richardson,P.
TITLE       Genetic and genomic tools for Xenopus research: The NIH Xenopus
            initiative
JOURNAL     Dev. Dyn. 225 (4), 384-391 (2002)
PUBMED      12454917
REFERENCE   2  (bases 1 to 1064)
AUTHORS     Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE       Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED      12477932
REFERENCE   3  (bases 1 to 1064)
AUTHORS     Klein,S. and Gerhard,D.S.
TITLE       Direct Submission
JOURNAL     Submitted (07-JUN-2004) National Institutes of Health, Xenopus Gene
            Collection (XGC), National Institute of Child Health and Human
            Development, 6100 Executive Boulevard, Room 4B01, Rockville, MD
            20892-7510, USA
REMARK      NIH-MGC Project
COMMENT     Contact: XGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Igor Dawid
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
```

info@bcgsc.bc.ca
Steve Jones, Sarah Barber, Mabel Brown-John, Yaron Butterfield,
Andy Chan, Steve S. Chand, William Chow, Alison Cloutier, Ruth
Featherstone, Malachi Griffith, Obi Griffith, Ran Guin, Nancy Liao,
Kim MacDonald, Amara Masson, Mike R. Mayo, Josh Moran, Ryan Morin,
Teika Olson, Diana Palmquist, Anca Petrescu, Anna Liisa Prahbu,
Parvaneh Saeedi, JR Santos, Angelique Schnerch, Ursula Skalska,
Duane Smailus, Jeff Stott, Miranda Tsai, George Yang, Jacquie
Schein, Asim Siddiqui, Rob Holt, Marco Marra.

```
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 152 Row: d Column: 10
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein.
FEATURES             Location/Qualifiers
     source          1..1064
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="MGC:82648 IMAGE:5049322"
                     /tissue_type="Ovary, Xenopus"
                     /clone_lib="NICHD_XGC_Ov1"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            1..1064
                     /gene="MGC82648"
                     /db_xref="LocusID:444046"
     CDS             35..742
                     /gene="MGC82648"
                     /codon_start=1
                     /product="MGC82648 protein"
                     /protein_id="AAH73035.1"
                     /db_xref="GI:49257301"
                     /db_xref="LocusID:444046"
                     /translation="MASLDRVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVD
                     VRLHEYRGTPEEKTYYTELWDVGGSVSASSVKSTRAVFYNAVNGIILVHDLTNKKS
                     SQNLYRWSLEALNRDLQPMGVLVTNGDYDREQFADNQIPLLVIGTKLDQIPEAKRNEV
                     LTRTAFLABDFNAEEINLDCTNTRCLAAGSNAVKLSRFDKVIRKRYPREGNLIPGF
                     SDRKRFGGGNFKSLHYD"
ORIGIN
Query Match              58.8%;   Score 418.2;    DB 5;   Length 1064;
Best Local Similarity    75.2%;   Pred. No. 8.1e-112;
Matches    535;  Conservative    0;  Mismatches  173;  Indels   3;  Gaps   1;

Qy    1 ATGGGCTCCCGGATCGGCTGAAGGTACTGCGTGTTGGGAGACTCAGTGTTGGGAAATCT 60
          |||||| |||||||||||| ||||||| ||||||  | ||||||||||| ||||||||
Db   35 ATGGGCTTCTGGACAGAGTGAAGGTTCTGGTGTTGGGGATTCAGGAGTGCGGGAGAGTCT 94

Qy   61 TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGACTGTGGGC 120
          |||| ||||||||  ||||||| |||||||||||| |||||||||| ||| ||||||||
Db   95 TCGTTGGTGCATCTGTCTGTGTGTCAGAATCAGGTTTGTGGGAACCCCTCGTGACAGTTGGC 154

Qy  121 TGCTCAGTGGATGTCAGAGTTCATGATTACAAGAAGAGGAACCCCAGAAGAGACCTAC 180
          ||||||||||||||| |||||||||||||||||  |||| | |||||||| || ||||
Db  155 TGCTCAGTGGATGTGCCGGCTCCACGAGTACAGGGACGAGGGACTCCAGAGCAGCAGGAAACCTAT 214

Qy  181 TACATAGAATTATGGATGTTGGAGGTCTGGGAGCTGAGGCTCTGGGAGCGTGAAAAGCACA 240
          ||||| ||||| |||| || ||||| ||||||||||| |||  | ||| | |||||||||
Db  215 TACACTGAGCTGTGGGACGTGGGTGGGTCCGTTGGAGTGCCAGCAGTGTGAAGAGCACC 274

Qy  241 AGAGCAGTATTCTACAACTCCGTAAATGTATTATTTTCGTACACGACTTAACAAATAAG 300
          |||||| ||| |||||| ||| || ||| |||  || ||| ||||| |||| | ||||
Db  275 AGGGCTGTGTTTTATAATGCAGTAAATGGGATCATCCTTGTCATGATCTTGACCAAATAAG 334

Qy  301 AAGTCCTCCCAAAACTTGCCTGCGTTCGTTGGTCATTGGAAGCTCTCAACACGGATTTGGTGCCA 360
          ||| |||| | |||| |||  || | | ||||||  ||||||| | ||| | ||| ||||||
Db  335 AAATCATCTCAGAACCTTTATCGCTGCGTCATTAGAGACCCCGAGATCTGCAGCCA 394

Qy  361 ACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATA 420
          | |||||||||||||  ||||| |||| ||| || ||| ||||| |||| | ||||||
```

```
Db    395 ATGGGAGTGCTAGTCACTAATGGGACTACGACAGAGAGCAATTTGCTGATAATCAGATC  454
           |||||||||||||||||||||||||||||||||||||||||||| |||| ||| |||
Qy    421 CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTT  480
           |||||||||||||||||||||||||||||||||||||||| ||||  |||| |||||||
Db    455 CCATTGTTGGTCATTGGAACCAAACTGGACCAAATTCCAGAGGCCGAAGCGCAATGAGGTC  514

Qy    481 TTAACTAGGAGCTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGC  540
           |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    515 CTGACACGAACTGCCTTTCTAGCTGAGGATTTAAATGCTGAAGAGATTAATCTTGACTGT  574

Qy    541 ACAAATCCACGGTACTTAGCTGCCAGGTTCTTCCAATGCTGTCAAGCTCAGTAGGTTTTT  600
           ||||  |||  ||| ||||||||||||||||||||||||||||||||||||||||||||
Db    575 ACGAACACACGATGCCTGGCTGCTGGGTTCTTCCAATGCTGTCAAACTCAGCCAGGTTTTT  634

Qy    601 GATAAGGTCATAGAGAGAGAGATACTTTTTAAGAGAAGGTAATCAGATTCCAGGCTTTCCT  660
           ||| ||||||| |||||||||||  ||||||||||||||||||||||||||||||||
Db    635 GACAAGGTTATTGAGAAAAGATAC---CCACGGGAAGGAAGGGAAATCTGATCCCTGGTTTCTCC  691

Qy    661 GATCGGAAAAGATTTGGGCAGGAGGACATTAAAGAGCCTTCATTATGACTGA  711
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    692 GATCGGAAACGCCTTTGGAGGGGGGGGACTTTAAGAGTCTCCACTATGACTGA  742


RESULT 11
BC078191
LOCUS          BC078191        1040 bp     mRNA        linear      VRT 26-JUL-2004
DEFINITION     Danio rerio cDNA clone MGC:100876 IMAGE:7141370, complete cds.
ACCESSION      BC078191
VERSION        BC078191.1   GI:50603855
KEYWORDS       MGC.
SOURCE         Danio rerio (zebrafish)
ORGANISM       Danio rerio
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
               Cypriniformes; Cyprinidae; Danio.
REFERENCE      1  (bases 1 to 1040)
AUTHORS        Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
               Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
               Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
               Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
               Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
               Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
               Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
               Carninci,P., Prange,C., Raha,S.S., Loqueliano,N.A., Peters,G.J.,
               Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
               McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
               Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
               Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
               Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
               Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
               Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
               Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
               Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
               Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE          Generation and initial analysis of more than 15,000 full-length
               human and mouse cDNA sequences
JOURNAL        Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED         12477932
REFERENCE      2  (bases 1 to 1040)
AUTHORS        Strausberg,R.
TITLE          Direct Submission
JOURNAL        Submitted (19-JUL-2004) National Institutes of Health, Mammalian
               Gene Collection (MGC), Cancer Genomics Office, National Cancer
               Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
               USA
REMARK         NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT        Contact: MGC help desk
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: Genome Institute of Singapore
               cDNA Library Preparation: S. Mathavan, Chia-Lin Wei, and Yijun
               Ruan, Genome Institute of Singapore
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               DNA Sequencing by: National Institutes of Health Intramural
               Sequencing Center (NISC),
               Gaithersburg, Maryland;
               Web site: http://www.nisc.nih.gov
               Contact: nisc mgc@hgri.nih.gov
               Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
               Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S.,
               Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
               Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R.,
               Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C.,
               McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W.,
               Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
               Young,A., Zhang,L.-H. and Green,E.D.

               Clone distribution: MGC clone distribution information can be found
               through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
               Series: IRAL Plate: 60 Row: e Column: 9
               This clone was selected for full length sequencing because it
               passed the following selection criteria: Hexamer frequency ORF
               analysis, Similarity but not identity to protein.
FEATURES             Location/Qualifiers
     source          1..1040
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /strain="Singapore local strain"
                     /db_xref="taxon:7955"
                     /clone="MGC:100876 IMAGE:7141370"
                     /tissue_type="Embryo, 7 different stages (from just
                     fertilized embryos to 72 hours just hatched baby fish)"
                     /clone_lib="GISZF001_ra"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
     CDS             42..743
                     /codon_start=1
                     /product="Unknown (protein for MGC:100876)"
                     /protein_id="AAH78191.1"
                     /db_xref="GI:50603856"
                     /translation="MASLDRVKVLVLGDSGVGKSSLVHLLCQNQVLGNPSWTVGCSVD
                     VRVHDYREGTPEEKAPYIEILWDVGGSVGSASSVKSTRAVPYNSVNGIILVHDLTNKKS
                     SQNLYRWSLEALSKDSSPTGIIVSNGYDRRQFAENAVPLLLIGTKFDQIPENKRNDV
                     LTRTAFLSEDFNABEINLDCTNPRXLAAGSNAVKLSRFFDKVIEKRYFTRDPSOMQS
                     FTDRRRFNFKSLHSD"
ORIGIN
Query Match            55.8%;    Score 396.6;    DB 5;     Length 1040;
Best Local Similarity  74.2%;    Pred. No. 1.9e-105;
Matches      501;  Conservative    0;  Mismatches 174;  Indels   0;  Gaps    0;

Qy    1 ATGGGCGTCCCTGATCGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCT  60
         ||||| ||||| ||||||||||||||||||||||||||||||||||| ||||||||
Db   42 ATGGGCATCTTTGGACAGGGTGAAAGTGCTGGTTTGGGAGAATTCTGGAGTAGGGAAGTCC  101

Qy   61 TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTCGGAAAATCCATCATGACTGTGGGC  120
         | | |||  ||  || |||||| |||||||||||||||||||||||||||| || ||||
Db  102 TCTCTTGTACACCTTCTTGCCCAGAATCAGGTTTTTGGGGAATCCCTCATGGACTGTGGGC  161

Qy  121 TGCTCAGTGGATGTCAGAGTTCATGATTGGGACGTTCATGATTACAAGAAGAGAACCCCAGAGACCTAC  180
         ||||||||||| || | |||||| ||| |||||| | |||||| ||||||||||||||||||||| ||
Db  162 TGCTCAGTGGACGTCAGGGTTCACGACTACAGAGAAGGGCACTCCAGAGCAGCCCAGAGCAGGCTTTC  221

Qy  181 TACATAGAATTATGGGATGTTGGAGGGTCTCGTGGGCAGTGCCAGCGCTGAAAAGCACA  240
         ||||| ||||| ||||| |||||||| | |||||| |||||||||| ||||||||  ||
Db  222 TACATTGAACTCCTGGAAGTGTGGGAGGATCTCGTTCCCAGCCAGCAGTGTTAAAAGCACC  281

Qy  241 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGACTTAACACAAATAAG  300
         |||||||||| ||||||| ||  |||||||| ||||  |||| |||  ||||||||||||
Db  282 AGAGCTGGTTTTACAATTCAGTTAAGTGATATCATTTTTAGTTCACGATCTGACTCAACAAG  341

Qy  301 AAGTCCTCCCAAAACTTGCGTCGTTGGCGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCA  360
         ||||||| |||||| ||||||  ||  |||||||||||| |||||| ||||||| ||||| |
Db  342 AAGTCCTCCCAGAATCTATACCGCTGGTCATTTAGAGGCACTAAGCAACAAGGACTCCTCTCCA  401

Qy  361 ACTGGAGTCTTCGTGACAAATGGGGATTATGATCAAGAAGACAGTTTGCTGATAACCAAATA  420
         ||||| |||| |||| || ||||||  ||| |||| || ||||||| |||  ||| |||
```

```
Db   402 ACTGGCATCATCGTATCCAACGGTGATTACGATAGAGAACAGCTGT 461
         |||| |||| ||| ||  |||||||| || |||||||||| | ||
Qy   421 CCACTGTTCGGTAATAGGGACTAAACTGGACCCAGATTCATGAAACAAAGCGCCATGAAGTT 480
         || | |  | ||| || |||| |||  |  |  |||||| ||| ||||||| ||||| || |
Db   462 CCTCTTCTGCTAATTGGCACCAAATTTGATCAGATCCCGGAAAAACAAGAGGAACGATGTC 521

Qy   481 TTAACTAGGACTGCTTCTATCTATCGAAGACTTCAATGCGGAGAATCAACCTGGATTGT 540
         || | ||| ||||| || ||| || | ||| |||||| |||| |||  | |||| |||
Db   522 CTGACCCGAACTGCCTTTCTATCGCGAGGTTCTTCCAAGCTCTCAAGCTCAGTAGGTTTTT 581

Qy   541 ACAAATCCACCGGTATCTTGCTGCTGGTTCATCCAAATGCTGTCAAGCTCAGTAGGTTTTT 600
         ||||||||||||| ||| ||||||||  | ||||  |||||| |||||| ||||||||||
Db   582 ACAAATCCACCGGTATCTTGCTGCTGGTTAATCGATCCCAGATTCATCAGATCAGTTCTTT 641

Qy   601 GATAAGGTAATAGAGACAGAAGAGATACTTTTTAAGAGAGAGACCCAGAGCTTCCAGGCTTCCT 660
         ||||||||||||||| ||||||||| |||| || ||| || | |||| | |||| || |||||
Db   642 GATAAGGTAATAGAGAGAAAAGATACTTTTCACAGAGACCCTGCCAGATGCAGAGCTTCACA 701

Qy   661 GATCGGAAAAGATTT 675
         || ||| |||  |||
Db   702 GACAGGAGGCGCTTT 716
```

RESULT 12
AY391440
LOCUS   AY391440   1291 bp   mRNA   linear   VRT 07-OCT-2003
DEFINITION Danio rerio hypothetical protein MGC23920-like protein mRNA,
            complete cds.
ACCESSION   AY391440
VERSION   AY391440.1 GI:37362247
KEYWORDS   .
SOURCE   Danio rerio (zebrafish)
  ORGANISM Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 1291)
  AUTHORS   Song,H.D., Wu,X.Y., Sun,X.J., Zhou,Y., Liu,T.X., Deng,M.,
            Zhang,G.W., Sheng,Y., Chen,Y., Ruan,Z., Jiang,C.L., Fan,H.Y.,
            Zon,L.I., KanKi,J.P., Look,A.T. and Chen,Z.
  TITLE   Gene Expression Profiling in the Zebrafish Kidney Marrow Tissue
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 1291)
  AUTHORS   Song,H.D., Wu,X.Y., Sun,X.J., Zhou,Y., Liu,T.X., Deng,M.,
            Zhang,G.W., Sheng,Y., Chen,Y., Ruan,Z., Jiang,C.L., Fan,H.Y.,
            Zon,L.I., KanKi,J.P., Look,A.T. and Chen,Z.
  TITLE   Direct Submission
  JOURNAL   Submitted (15-SEP-2003) State Key Lab for Medical Genomics,
            Shanghai Institute of Hematology, Ruijin Hospital Affiliated to
            Shanghai Second Medical University, 197 Rui Jin Road II, Shanghai
            200025, P. R. China
FEATURES   Location/Qualifiers
     source   1..1291
                     /organism="Danio rerio"
                     /mol_type="mRNA"
                     /db_xref="taxon:7955"
                     /clone="RK025A2D04"
     CDS   3..704
                     /codon_start=1
                     /product="hypothetical protein MGC23920-like protein"
                     /protein_id="AAQ91252.1"
                     /db_xref="GI:37362248"
                     /translation="MASLDRVKVLVLGDSGVGKSSLVHLLCQNVLGNPSWTVGCSVD
                     VRVHDYRRGTPEKAFYIBLWDVGSGVSASSVKSTRAVFYNSVNGIILVHDLTNKKS
                     SQNLYRWSLEALSKDSSPTGIIVSNGDYDREQFAENAVPLLLIGTKPDQIPENKRNDV
                     LTRTAFLSEDFNAEEINDCTNPRYLAAGSSNAVKLSRFPDKVIEKRYFTRDPSQMQS
                     FTDRRRFNFKSLHSD"
ORIGIN

Query Match   55.6%; Score 395; DB 5; Length 1291;
Best Local Similarity 74.1%; Pred No. 5.6e-105;
Matches 500; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

RESULT 13
BD076273
LOCUS   BD076273   385 bp   DNA   linear   PAT 27-AUG-2002
DEFINITION 5' EST of tissue-nonspecific secretory protein.
ACCESSION   BD076273
VERSION   BD076273.1 GI:22621876
KEYWORDS   JP 2001512011-A/221.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 385)
  AUTHORS   Edwards,J.B.D.M., Duclert,A. and Lacroix,B.
  TITLE   5' EST of tissue-nonspecific secretory protein
  JOURNAL   Patent: JP 2001512011-A 221 21-AUG-2001;
            GENSET
COMMENT   OS   Homo sapiens (human)
            PN   JP 2001512011-A/221
            PD   21-AUG-2001
            PF   31-JUL-1998 JP 2000505289
            PR   01-AUG-1997 US 08/905135
            PI   JEAN BAPTISTE DUMAS MILNE EDWARDS,AYMERIC DUCLERT,BRUNO PI
            LACROIX
            PC   C12N15/09,C12N15/09,C07K14/47,C12Q1/68,C12N15/00,C12N15/00 CC

```
Qy     1 ATGGGCTCCCTGGATCGGGTGAAGGTACTGTGTTGGGAGACTCAGTGTGTTGGGAAATCT 60
         ||||||| || |||| |||||||| |||||| || |||||| || |  | |||||| ||
Db     3 ATGGGCATCTTTGGACAGGGTGAAAGTGCTGGTTTTGGGAGATTCTGGAGTAGGGAAGTCC 62

Qy    61 TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGACTGTGTGGGC 120
         || | ||| ||| ||| |||||||||||| |||  |||||||| ||||| |||||||||||
Db    63 TCTCTTGTACACCTTCTTTGCCAGAATCAGGTTTTGGGAATCCCTCATGGACTGTGTGGGC 122

Qy   121 TGCTCAGTGGATGTCAGAGTTCATGATGAAGAGGAACCCCAGAAGAGAAGAGAAGACCTAC 180
         |||||||||| ||||| ||||||| | ||||||||| || ||| ||||||||||||| |||
Db   123 TGCTCAGTGGACGTCAGGGTTCACGACTACAGAAGGCACTCCAGAGAAGAGAAGAGGCTTTC 182

Qy   181 TACATAGAATTATGGGATGTTGGAGGTCTGTGGGACGTCGTGGGCAGTGCCAGCAGTGTTA 240
         |||||||| || ||||| |||||||| ||||||||||||||||||||| || ||||| |
Db   183 TACATTGAACTCGGGATGTTGGAGGATCTGTGGGACGTCGTGGGCAGTGGTGTTAAAAGCACC 242

Qy   241 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGACTTAACAAATAAG 300
         |||||||| | ||||||||| || ||||||||| |||| | |||| ||||| |||||||
Db   243 AGAGCTGTGTTTACAATTCAGTTAATGGTATCATTTTAGTTGGTATCATTTTAGTTCGACTAACAAG 302

Qy   301 AAGTCCTCCAAAACTGGCGTGGTCGTTGGTGTCATTGGGGATTCATTGGCTTCTCAACAGGGATTTGGTGCCA 360
         ||||| ||||||||||| ||| | ||||  || |||||| || || |||||||| |  ||||| |||||| ||
Db   303 AAGTCTTCCCAGAATCTATACCGCTGGTCACTAGAGGCACTCGAGCAAGAGACTTCAGACGATTGTGATAACCTCCA 362

Qy   361 ACTGGAGTCTTGGTGACAAATGGGATTATGATCCAAGACAGTTGCTGATTAACAAATA 420
         |||||| |  ||||| |||| |||| || ||||| |  |||||| ||||| ||  || |
Db   363 ACTGGCATCATCGTATCCAACGGTGATTACGATAGAGAACAGCTGTTGCAGAATGCTGTT 422

Qy   421 CCACTGTTGGTAATAGGGGACTAAACTGGACCCAGATTCATGAAACAAAGCGCCATGAAGTT 480
         || | |  | || || |||| || |||| ||||||| |||| |  | |||| |||||| ||
Db   423 CCTCTTCTGCTAATTGGCACCAAATTTGATCAGATCCCGGAAAAACAAGAGGAACGATGTC 482

Qy   481 TTAACTAGGACTGCTTCTATCTATCGAAGACTTCAATGCGGAGAATCAACCTGGACTGC 540
         || | ||| ||||| || ||| ||  |||| |||||| |||| |||||  |||| | ||
Db   483 CTGACCCGAACTGCCTTTCTATCGCGAGGATTTCAATCGGAGAATTAATTGGGCACCTGGATTGC 542

Qy   541 ACAAATCCACCGGTATCTTGCTGCTGGTTCATCCAAATGCTGTCAAGCTCAGTAGGTTTTT 600
         |||||||||||| ||| || ||||||  | ||   ||||| ||| || |||||||| || |
Db   543 ACAAATCCACCGGTATCTTGCTGCTGGTTAATCGATCAAATGCTGTTAAACTGTGTTCAGTCGATTCTTT 602

Qy   601 GATAAGGTCATAGAGAGATACTTTTTAAGAGAGAGACCCAGAGCTTCCAGGCTTCCT 660
         ||||||||  ||||||||| |||| |  ||| ||| |  ||| | |||| ||||| |
Db   603 GATAAGGTAATAGAGAAAAGATACTTCACAGAGACCCCTAGCCAGATGCAGAGCTTCACA 662

Qy   661 GATCGGAAAAGATTT 675
         || ||| ||| ||||
Db   663 GACAGGAGGCGCTTT 677
```

```
Qy    361 ACTGGAGTCTTGGTG 375
          |||||||||||||||
Db    371 ACTGGAGTCTTGGTG 385

RESULT 14
CQ719549
LOCUS       CQ719549                 223 bp   DNA        linear   PAT 03-FEB-2004
DEFINITION  Sequence 5483 from Patent WO02068579.
ACCESSION   CQ719549
VERSION     CQ719549.1  GI:42280406
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Venter,C.J., Adams,M.C., Li,P.W. and Myers,B.W.
  TITLE     Kits, such as nucleic acid arrays, comprising a majority of
            humanexons or transcripts, for detecting expression and other uses
            thereof
  JOURNAL   Patent: WO 02068579-A 5483 06-SEP-2002;
            PE Corporation (NY) (US)
FEATURES             Location/Qualifiers
     source          1..223
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match             31.4%;  Score 223;  DB 6;  Length 223;
Best Local Similarity   100.0%;  Pred. No. 2.4e-54;
Matches  223;  Conservative   0;  Mismatches   0;  Indels    0;  Gaps   0;

Qy    384 GGATTATGATCAAGAACAGTTTGCTGATAACCAATCCACTGTTGGTAATAGGGACTAA 443
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGATTATGATCAAGAACAGTTTGCTGATAACCAATCCACTGTTGGTAATAGGGACTAA 60

Qy    444 ACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCTGGC 503
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCTGGC 120

Qy    504 TGAGGATTTCAATCCAGAAGAGAAATTAATTTGGACTGCACAAATCCACGTACTTAGCTGC 563
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGAGGATTTCAATCCAGAAGAGAAATTAATTTGGACTGCACAAATCCACGTACTTAGCTGC 180

Qy    564 AGGTTCTTCCAATGCTGTCAAGCTCAGTAGGTTTTTTGATAAG 606
          |||||||||||||||||||||||||||||||||||||||||||
Db    181 AGGTTCTTCCAATGCTGTCAAGCTCAGTAGGTTTTTTGATAAG 223

RESULT 15
AC133474
LOCUS       AC133474               65838 bp   DNA        linear   PRI 03-DEC-2002
DEFINITION  Homo sapiens 3 BAC RP11-719N16 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC133474
VERSION     AC133474.2  GI:23334664
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 65838)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            AlBrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,K.L., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
```

```
           blastn
CC         identity 97
CC         region 1..347
CC         id AA023764
CC         est
CC         blastn
CC         identity 95
CC         region 145..384
CC         id C03036
CC         est
CC         blastn
CC         identity 98
CC         region 2..71
CC         id C03036
CC         est
CC         blastn
CC         identity 99
CC         region 1..193
CC         id R08519
CC         est
CC         blastn
CC         identity 94
CC         region 193..263
CC         id R08519
CC         est
CC         Von Heijne matrix
CC         score 4.8
CC         seq SLVHLLCQNQVLG/NP
CC         n=a, g, c or t
FH   Key             Location/Qualifiers
FT   misc_feature    39..385
FT   misc_feature    146..385
FT   misc_feature    11..80
FT   misc_feature    39..231
FT   sig_peptide     232..302
FT   misc_feature    11..109
FT   misc_feature    287
FT   misc_feature    331..
FEATURES             Location/Qualifiers
     source          1..385
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match             51.4%;  Score 365.8;  DB 6;  Length 385;
Best Local Similarity   97.3%;  Pred. No. 2.1e-96;
Matches  365;  Conservative   5;  Mismatches   5;  Indels    0;  Gaps   0;

Qy      1 ATGGGCTCCCTGGATCGGGTGAAGGTACTGGGTGTTGGGAGACTCAGGTGTTGGGAAATCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     11 ATGGGCTCCCTGGATCGGGTGAAGGTACTGGGTGTTGGGAGACTCAGGTGTTGGGAAATCT 70

Qy     61 TCGTTAGTCCATCTCCTATGCCAAATCAAGTGCTGGAAATCCATCATGACTGTGGGGC 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     71 TCGTTAGTCCATCTCCTATGCCAAATCAAGTGCTGGAAATCCATCATGACTGTGGGGC 130

Qy    121 TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAAGGAACCCCCAGAAGAGAAGACCTAC 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    131 TGCTCAGTGGATGTCAGAGTKCATGATTACAAAGAAGAAGGAACCCCCAGAAGAGAAGACCTAC 190

Qy    181 TACATAGAATTATGGGATGTTGGAGGCTCTGGGCAGTGGCCAGTGGCCGTGAAAAGCACA 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    191 TACATAGAATTATGGGATGTTGGAGGCTCTGGGCAGTGGCCAGTGGCCGTGAAAAGCACA 250

Qy    241 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTTCGTACACGACTTAACAAATAAG 300
          ||||||||||||||||||||||||||||||||||||||:: |||||||||||
Db    251 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATWNYCGTACACGACTTAACACVSATGGG 310

Qy    301 AAGTCCTCCCAAAACTTGCCTCGTTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCA 360
          |||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db    311 AAGTCCTCCCAAAMTTGCGNCGTTGGTCATTGGAAGCTCTCAACAGGGGATTTGGTGCCA 370
```

continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht
ml.

```
FEATURES             Location/Qualifiers
source               1..65838
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-719N16"
misc_feature         1..2005
                     /note="overlaps bases 117724..119726 of clone AC117472"
                     /function="clone overlap"
repeat_region        complement(1..420)
                     /rpt_family="L2"
STS                  2133..2353
                     /standard_name="SHGC-68332"
repeat_region        2818..3009
                     /rpt_family="MIR"
repeat_region        complement(4474..5354)
                     /rpt_family="L1M4"
repeat_region        5358..5382
                     /rpt_family="(TTTTG)n"
repeat_region        complement(5383..5642)
                     /rpt_family="AluSx"
repeat_region        complement(5643..7192)
                     /rpt_family="L1M4"
repeat_region        7550..7851
                     /rpt_family="AluY"
repeat_region        7892..8199
                     /rpt_family="MER1B"
repeat_region        complement(8480..8586)
                     /rpt_family="MER5A"
repeat_region        9046..9175
                     /rpt_family="L1ME"
repeat_region        complement(9498..10116)
                     /rpt_family="L1MC5"
repeat_region        11023..11046
                     /rpt_family="AT_rich"
repeat_region        complement(12233..12522)
                     /rpt_family="AluSg"
repeat_region        complement(12953..13127)
                     /rpt_family="MIR"
repeat_region        16272..16299
                     /rpt_family="MIR"
repeat_region        complement(13247..13436)
                     /rpt_family="MER5A"
repeat_region        13772..13813
                     /rpt_family="AT_rich"
repeat_region        17909..18207
                     /rpt_family="AluY"
repeat_region        complement(18719..19081)
                     /rpt_family="MER47A"
STS                  20760..20949
                     /standard_name="SHGC-77253"
```

          Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
          Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
          Earnhart,C., Edgar,D., Edwards,C.C.C., Elhaj,C., Escotto,M.,
          Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
          Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
          Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
          Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
          Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
          Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
          Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
          Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
          Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
          Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
          Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
          Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
          Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,K.,
          Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
          Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
          Nelson,D., Newton,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
          Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
          Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
          Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,V., Rives,M.,
          Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
          Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
          Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
          Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
          Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
          Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
          Warren,R., Washington,C., Watlington,S., Williams,G.,
          Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,
          Wu,Y.P., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
          Gibbs,R.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 65838)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (13-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 65838)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (27-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4 (bases 1 to 65838)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (03-DEC-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Sep 27, 2002 this sequence version replaced gi:22830213.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
            STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
            Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
            Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < le-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence

```
repeat_region   21591..21610
                /rpt_family="(TG)n"
repeat_region   22278..22315
                /rpt_family="AT_rich"
repeat_region   22316..22596
                /rpt_family="AluSx"
repeat_region   22597..22634
                /rpt_family="(TAAA)n"
repeat_region   22957..23107
                /rpt_family="AluJb"
repeat_region   23327..23425
                /rpt_family="LTR29B"
repeat_region   23401..23779
                /rpt_family="MER39"
repeat_region   complement(24441..24627)
                /rpt_family="MER5A"
repeat_region   complement(25067..25188)
                /rpt_family="HAL1"
repeat_region   25240..25367
                /rpt_family="L2"
repeat_region   25508..25664
                /rpt_family="AluJb"
repeat_region   25670..25958
                /rpt_family="AluJo"
repeat_region   25959..25985
                /rpt_family="AT_rich"
repeat_region   complement(26771..26936)
                /rpt_family="MLT1H"
repeat_region   complement(27115..27217)
                /rpt_family="MLT1H"
repeat_region   complement(28823..29001)
                /rpt_family="MIR"
repeat_region   complement(29469..29599)
                /rpt_family="L2"


Query Match        21.7%;   Score 154.2;  DB 9;   Length 65838;
Best Local Similarity  98.1%;   Pred. No. 5.8e-34;
Matches  156;  Conservative   0;  Mismatches   3;   Indels   0;   Gaps   0;

Qy     383 GGGATTATGATCAAGAACAGTTGCTGATAACCAAATACCACTGTTGGTAATAGGGACTA 442
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11832 GGGATTATGATCAAGAACAGTTGCTGATAACCAAATACCACTGTTGGTAATAGGGACTA 11891

Qy     443 AACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTTTTAACTAGGACTGCTTTCCTGG 502
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   11892 AACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTTTTAACTAGGACTGCTTTCCTGG 11951

Qy     503 CTGAGGATTTCAATCCAGAAGAAAATTAATTGGACTGCA 541
           |||||||||||||||||||||||||||       ||  |
Db   11952 CTGAGGATTTCAATCCAGAAGAAAATTAATTTGGTATGTA 11990


Search completed: June 30, 2005, 23:29:36
Job time : 3261.89 secs
```

```
                    GenCore version 5.1.6
            Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:     June 30, 2005, 13:44:48 ; Search time 133.331 Seconds
                                      (without alignments)
                                      8725.600 Million cell updates/sec

Title:      US-09-945-173-3
Perfect score: 711
Sequence:   1 atggcgtccctgatcggt.........agagcctcattatgactga 711

Scoring table: IDENTITY NUC
               Gapop 10.0 , Gapext 1.0

Searched:   1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters:    2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued Patents NA:*
             1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
             2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
             3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
             4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
             5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
             6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	66.4	9.3	1087	4	US-09-270-767-13908	Sequence 13908, A
2	41.6	5.9	1141	4	US-09-806-708B-22	Sequence 22, Appl
c 3	36.8	5.2	7218	1	US-08-232-463-14	Sequence 14, Appl
c 4	35.2	5.0	1141	4	US-09-806-708B-22	Sequence 22, Appl
5	34.8	4.9	1212	2	US-09-092-770-18	Sequence 18, Appl
6	34.8	4.9	1212	3	US-09-222-851-18	Sequence 18, Appl
7	34.4	4.8	1182	4	US-09-248-796A-779	Sequence 779, App
8	34.2	4.8	810	4	US-09-248-796A-6393	Sequence 6393, Ap
9	34	4.8	735	4	US-09-270-767-13913	Sequence 13913, A
10	33.8	4.8	2742	4	US-09-134-000C-2635	Sequence 2635, Ap
11	33.8	4.8	45090	4	US-09-949-016-12690	Sequence 12690, A
12	33.8	4.8	45091	4	US-09-949-016-16195	Sequence 16195, A
13	33.2	4.7	1008	4	US-09-134-000C-2379	Sequence 2379, Ap
14	33	4.6	121049	4	US-09-949-016-17513	Sequence 17513, A
c 15	32.8	4.6	474	4	US-09-621-976-18033	Sequence 18033, A
16	32.8	4.6	618	4	US-09-270-767-28849	Sequence 28849, A
17	32.8	4.6	709	4	US-09-270-767-12973	Sequence 12973, A
18	32.6	4.6	126254	4	US-09-949-016-15341	Sequence 15341, A
19	32.6	4.6	786431	4	US-09-751-389-3	Sequence 3, Appli
20	32.4	4.6	601	4	US-09-949-016-81122	Sequence 81122, A
c 21	32.4	4.6	14566	4	US-09-949-016-16765	Sequence 16765, A
c 22	32.4	4.6	48994	4	US-09-949-016-14091	Sequence 14091, A
c 23	32.4	4.6	109025	4	US-09-949-016-12609	Sequence 12609, A
c 24	32.4	4.6	109025	4	US-09-949-016-17567	Sequence 17567, A
25	32.2	4.5	6508	4	US-09-995-917A-2	Sequence 2, Appli
c 26	32.2	4.5	80717	4	US-09-949-016-14968	Sequence 14968, A
c 27	32.2	4.5	155517	4	US-09-949-016-16191	Sequence 16191, A
c 28	32.2	4.5	580073	4	US-08-545-528D-1	Sequence 1, Appli
c 29	32	4.5	601	4	US-09-949-016-139959	Sequence 139959,
c 30	32	4.5	148567	4	US-09-801-876B-3	Sequence 3, Appli
c 31	32	4.5	148567	4	US-10-254-869-3	Sequence 3, Appli
c 32	32	4.5	148567	4	US-10-667-442-3	Sequence 3, Appli
33	31.8	4.5	312474	4	US-09-949-016-17434	Sequence 17434, A
34	31.8	4.5	1164	4	US-09-710-279-3179	Sequence 3179, Ap
35	31.8	4.5	3603	4	US-09-710-279-3419	Sequence 3419, Ap
c 36	31.6	4.5	134890	4	US-09-949-016-15602	Sequence 15602, A
37	31.6	4.4	601	4	US-09-949-016-194092	Sequence 194092,
38	31.6	4.4	1173	4	US-09-248-796A-6337	Sequence 6337, Ap
39	31.6	4.4	1543	3	US-09-364-230-17	Sequence 17, Appl
40	31.6	4.4	17922	4	US-09-949-016-15018	Sequence 15018, A
c 41	31.6	4.4	80632	4	US-09-949-016-12951	Sequence 12951, A
c 42	31.6	4.4	137949	4	US-09-949-016-12196	Sequence 12196, A
c 43	31.6	4.4	137956	4	US-09-949-016-17260	Sequence 17260, A
c 44	31.6	4.4	202001	4	US-09-734-674-3	Sequence 3, Appli
c 45	31.4	4.4	601	4	US-09-949-016-105459	Sequence 105459,

ALIGNMENTS

```
RESULT 1
US-09-270-767-13908
; Sequence 13908, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
;  APPLICANT: Homburger et al.
;  TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
;  FILE REFERENCE: File Reference: 7326-094
;  CURRENT APPLICATION NUMBER: US/09/270,767
;  CURRENT FILING DATE: 1999-03-17
;  NUMBER OF SEQ ID NOS: 62517
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13908
;  LENGTH: 1087
;  TYPE: DNA
;  ORGANISM: Drosophila melanogaster
US-09-270-767-13908

Query Match             9.3%;  Score 66.4;  DB 4;  Length 1087;
Best Local Similarity   52.0%;  Pred. No. 3.8e-12;
Matches  182;  Conservative   0;  Mismatches  156;  Indels  12;  Gaps   1;

Qy      1 ATGGCGTCCCTGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCT  60
            |||||||       ||||    ||| |||||||||  |||||||||||||||||| |||||
Db    269 ATGGCGGATGAATAATCGAGTGCGAATTGTTGTGTTGGGCGATTCGGGTGTGGGGAAGACC 328

Qy     61 TCGTTAGTCCATCTTCCTATGCCAAAATCAAGTGCTGGGAAATCATCATGGACTGTGGGC 120
            |||||  ||||||  ||   |||   |||| |||| |||| || ||||||||||||||||
Db    329 TCCCTGACGCACCTGATCACCCACAACGAAGCCCTCATCCGGCCCGGCTGGACGGTGGGC 388

Qy    121 TGCTCAGTGGATGTCAGAGTTCATGATTCATGATTACAAAGAAGGAACCCCAGAAGAGACCTAC 180
            |||| ||||| |||  ||||  ||| |   |||||||||  |||||| || ||||||| |||||
Db    389 TGCAACATCCAGGTGAAGATGCATCCGTTCAGGGAGGCGCACCGCTCGCGAGTGTCCCTAC 448

Qy    181 TACATAGAATTATGGGATGTTGGAGGCTTCGTGTGGGACGGTG----------AACCAAGAAGCACA 240
            |||     ||||      |||| |||||  ||||||||   ||       |||||||||||||||
Db    449 TTCGTGGAGCTGTTCGATGTTGGGGGATCGGCTG--------------AACCACAGAACACG 496

Qy    241 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACGACTTAACAAATAAG 300
            ||||||  |||||| ||||| |||||||| ||||  ||||||||  |||||| ||||
Db    497 CGCAGCGTCTTCATACGGGCATCGATCGCATGGAATCATTCTGGTGCACGACCTTACCAACGCC 556

Qy    301 AAGTCCTCCCAAAACTTGCCTCGTCGTTGGTCATTGGAAGCTCTCAACAGGA 350
            ||||       |||||| |||     ||   ||||||||  |||| |||||||
Db    557 AAGTCGCAGAGGCAGCTAATCGACTGGCGTGACGAGATCGTCAACAAGGA 606

RESULT 2
US-09-806-708B-22
; Sequence 22, Application US/0906708B
```

```
;  Patent No. 6784342
; GENERAL INFORMATION:
;  APPLICANT: The University of British Columbia
;  TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
;  FILE REFERENCE: 4810-58741
;  CURRENT APPLICATION NUMBER: US/09/806,708B
;  CURRENT FILING DATE: 2001-04-03
;  PRIOR APPLICATION NUMBER: US 60/147,133
;  PRIOR FILING DATE: 1999-08-04
;  NUMBER OF SEQ ID NOS: 23
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
;  LENGTH: 1141
;  TYPE: DNA
;  ORGANISM: Artificial sequence
;  FEATURE:
;  NAME/KEY: promoter
;  LOCATION: (1)..(1141)
;  OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
; US-09-806-708B-22

Query Match          5.9%;  Score 41.6;  DB 4;  Length 1141;
Best Local Similarity  12.2%;  Pred. No. 0.0017;
Matches  63;  Conservative  188;  Mismatches  264;  Indels  1;  Gaps  1;

Qy  153 AGAAGGAACCCAGAAGAGAAGAGACCTACTACATAGAAATTATGGGATGTTGGAGGCTCTGT 212
Db  252 MVAWBTAHRRRYRNNGWTBAMAYRRRWTMNNNNNAKAMCKRAKYWGWNRABVNSTCTTWKS 311

Qy  213 GGGCAGTGCCAGCAGCGTGAAAAGCACAAGAGACCAGTATTCTACAACTCCGTAAATGGTAT 272
Db  312 KTTKVRTSCWANNCRAGDANKDHKWWKSWSAAMGVYWNNNNNNNWTYKKARHBARWDWVWH 371

Qy  273 TATTTTCGTACACGACTTAACAAATAAGAGAGTCCTCCCAAAACTTGCGTCGTTGGTCATT 332
Db  372 SAWKKWHANIAAHYSRKKWTBYKRKTMVNNNNGTTMWKRMWAWYWKMDMDWBGTYNNNNNG 431

Qy  333 GGAAGCTCTCAACAGGGGATTTGGTGCCAACTGGAGTCTTGGTGACAAATGGGGATTATGA 392
Db  432 GRTYGWTKNKGMWTYYKWKANNCKWRAWDHKTCTHNNTTWWKMKTYWNNCYWKSMTNGK 491

Qy  393 TCAAGAACAGTTTGCTGATAACCAAATACCACCTGTTGGTAATAGGGACTAAACTGGACCA 452
Db  492 SHRBAAAVVTWYMWWWRRYAHANNNWDYWWKACTWYKYBVCSKWWNNYAAWYTKSSWNY 551

Qy  453 GATTCATGAAACAAAGCGCCATGAAGTTTTAACTAGGACTTGCTTTCCTGCTGAGGATTT 512
Db  552 TSRYYRWKTNNSWRWRSDTRSMGRANNYARABHYGYKWNTRWWBWSHTWBHRAGAAAHYW 611

Qy  513 CAATCCAGAAGAGAATTAATTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCTTC 572
Db  612 MBMMYBAKCHCMKAWYKAKKYAGAGGSNNNNNNNNNNNATCARDDYYAASRWYAM 671

Qy  573 CAATGCTGTCAAGCTCAGTAGGTTTTTTGATAAGGTCATAGAGAAGAGATACTTTTTAAG 632
Db  672 ANAKWYYYKBAANNAYYTHANNWWGCWNNATDTRRTMWK-NNNNNNAGTWKNNNNNNAKN 730

Qy  633 AGAAGGTAATCAGATTCCAGGCTTTCCTGATCGGAA 668
Db  731 ASAAKNYAAAAVKAAKKHWRWANKWAMRGWHADARA 766
```

```
RESULT 3
US-08-232-463-14/c
; Sequence 14, Application US/08232463
;  Patent No. 5670367
; GENERAL INFORMATION:
;  APPLICANT: DORNER, F.
;  APPLICANT: SCHEIFLINGER, F.
;  APPLICANT: FALKNER, F. G.
;  TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
;  NUMBER OF SEQUENCES: 52
;  CORRESPONDENCE ADDRESS:
```

```
;  ADDRESSEE: Foley & Lardner
;  STREET: 1800 Diagonal Road, Suite 500
;  CITY: Alexandria
;  STATE: VA
;  COUNTRY: USA
;  ZIP: 22313-0299
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: IBM PC compatible
;  OPERATING SYSTEM: PC-DOS/MS-DOS
;  SOFTWARE: PatentIn Release #1.0, Version #1.25
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/232,463
;  FILING DATE:
;  CLASSIFICATION: 435
;  PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: US/07/935,313
;  FILING DATE:
;  APPLICATION NUMBER: EP 91 114 300.6
;  FILING DATE: 26-AUG-1991
;  ATTORNEY/AGENT INFORMATION:
;  NAME: BENT, Stephen A.
;  REGISTRATION NUMBER: 29,768
;  REFERENCE/DOCKET NUMBER: 30472/114 IMMU
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE: (703)836-9300
;  TELEFAX: (703)683-4109
;  TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 7218 base pairs
;  TYPE: nucleic acid
;  STRANDEDNESS: single
;  TOPOLOGY: linear
;  IMMEDIATE SOURCE:
;  CLONE: pTZgpt-F1s
; US-08-232-463-14

Query Match          5.2%;  Score 36.8;  DB 1;  Length 7218;
Best Local Similarity  3.5%;  Pred. No. 0.26;
Matches  14;  Conservative  210;  Mismatches  172;  Indels  0;  Gaps  0;

Qy    99 AAATCCATCATGGACTGTGGGCTGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGG 158
Db  1457 AAAGAGATAGAAGAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1398

Qy   159 AACCCCAGAAGAGAAGAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAG 218
Db  1397 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1338

Qy   219 TGCCAGCGTGAAAAGCACAAGAGACCAGTATTCTACAACTCCGTAAATGGTATTATTTTT 278
Db  1337 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1278

Qy   279 CGTACACGACTTAACAAATAAGAAGAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGC 338
Db  1277 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1218

Qy   339 TCTCAACAGGGATTTGGTGCCAACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGA 398
Db  1217 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1158

Qy   399 ACAGTTTGCTGATAACCAAATACCACCTGTTGGTAATAGGGACTAAACTGGACCAGATTCA 458
Db  1157 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1098

Qy   459 TGAAACAAAGCGCCATGAAGTTTTAACTAGGACTGC 494
Db  1097 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRATCGC 1062
```

```
RESULT 4
US-09-806-708B-22/c
```

; Sequence 22, Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. FAE1 promoters
US-09-806-708B-22

Query Match   5.0%;   Score 35.2; DB 4; Length 1141;
Best Local Similarity   10.6%;   Pred. No. 0.28%;
Matches   59;   Conservative 184;   Mismatches 311;   Indels   0;   Gaps   0;

Qy   151 AAAGAAGGAACCCCAGAAGAGAAGAAGACCTACTACATAGAATTATGGGATGTTGGAGGCTCT 210
                ::: || ||:|| |: : |:| : || : ||: || |
Db   929 MAMKAKQNNNNNNAYWTACYNRAATNNKMATHWMKWTHGAHSKRRTRHHTRTCRRTKYNNN 870

Qy   211 GTGGGCAGTGCCAGCAGCGTGAAAAGCACAAGAGACAGTATTCTACAACTCCGTAAATGGT 270
                ::   :: |: :: ::: | | :: :: :: :: :
Db   869 NNNARTVYWYHHAARRWMNAWWTRTNNNNNNNACRNTRTWWABWKHSWCNNNNNNNNN 810

Qy   271 ATTATTTTCGTACACGACTAACAAATAAGAAGTCCTCCCAAAACTTGCCGTCGTTCGGTCA 330
                :: | | ::|| | || |:||| | | |: | || | || | :
Db   809 NNNNNNNTWCHYTTANABBCYRANNNNAAARMARTCNNYMHAAVTTTHTDWCYKTWMNTWY 750

Qy   331 TTGGAAGCTCTCAACAGGGATTGGTGCCAACTGGACGTCTTGGTGACAAATGGGGATTAT 390
                :: ||| : :||| || :: ||
Db   749 WDMMTTMBTTTTRNMTTSTNMTNNNNNMWAACTNNNNNNNMWKAYYAHATNNWGCWNNNTD 690

Qy   391 GATCAAGAACAGTTTGCTGATAACCAAATACCACCTGTTGGTAATAGGGACTAAAACTGGAC 450
                ::::: || |::::: :|:|||| ||: ::::::|
Db   689 ARRTRNNTTVMRRRWMNTKTRWYSTTRRHYTGATNNNNNNNNNNNNSCCTCTRMM 630

Qy   451 CAGATTCATGAAACAAAGCGCCAAGCTGAAGTTTTAACTAGGACTGCTTTCCTGGCTGAGGAT 510
                :::: |: || :|| ||: : |||::::|| :| |::
Db   629 TMRWTMKGDGMTVRKKVKWRDTTCTYVDVWADSWVWYANWMRCRDVTYTRNNTYCKSYA 570

Qy   511 TTCAATCCAGAGAGAAATTAATTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCT 570
                : ::::| :|||| |:| |: |::|| | ||::|||:|:: : ||::
Db   569 HSYWYWSNNAMWYRRYSARNWSSMARWTTRNNWWMSGBVRMRWAGTMWWRHWNNNNTDTR 510

Qy   571 TCCAATGCTGTGCAAGCTCAGTAGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTTA 630
                :|::||: |:|: ::::| |:|: |||||:: ::: |:|::
Db   509 YYWWWKRWARBTTTVDSMCNAKSMRGNNWRAMKMWRAMKMWWAAINDAGAMDHWTYWMGNNTMW 450

Qy   631 AGAGAAGGTAATCAGATTCCAGGCTTTCCTGATCGGAAAAGATTTGGGGCAGGAACATTA 690
                :::|:::|| ||| |:::||::|:|: | ||||:
Db   449 MRRAWKMMNNMAWCRRAYCCNNNNNRACVWHKGHKMWRWTWKYMWKAACNNNNBKAMYMRVA 390

Qy   691 AAGAGGCTTCATTA 704
                ||| :||| ::::
Db   389 WMMYSRDTTNTDWM 376

RESULT 5
US-09-092-770-18
; Sequence 18, Application US/09092770
; Patent No. 5973119
; GENERAL INFORMATION:
; APPLICANT: Coats, Steven R.

; APPLICANT: Bass, Michael B.
; APPLICANT: Robinson, Murray O.
; TITLE OF INVENTION: No. 5973119el Cyclin B Genes and Proteins
; FILE REFERENCE: A-524
; CURRENT APPLICATION NUMBER: US/09/092,770
; CURRENT FILING DATE: 1997-06-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Mouse
US-09-092-770-18

Query Match   4.9%;   Score 34.8; DB 2; Length 1212;
Best Local Similarity   29.3%;   Pred. No. 0.4;
Matches   79;   Conservative 40;   Mismatches 151;   Indels   0;   Gaps   0;

Qy   365 GAGTCTTGGTGACAAATGGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATACCAC 424
                | || || :   :   :| ||: || : | || ::
Db   512 GNTTYATGYTNACNCARAARGAYGTNAAYAARAAYATGYTNACARYTNARYTNGGNATHACNW 571

Qy   425 TGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTTTAA 484
                :: :   ::: |: |:: |||| :: |: || :|| ||:
Db   572 SNYTNTTYATHGCNWSNAARYTNGARGARATHTAYGCNCNAARYTNCARGARYTNCARGCACA 631

Qy   485 CTAGGACTGCTTTCCTGCTGAGGATTCAATCCAGAGAAATTAATTGGACTGCACAA 544
                :||||:| || : :: ||| ||::|| || :: |: |: |
Db   632 AYGTNACNGAYGGNGCNTGYWSNGARGTNGAYATHYTNAARATGAYYTNAAYATHYTNA 691

Qy   545 ATCCACGTACTTAGCTGCAGGTTCTTCCAATGCTGTGCAAGCTCAGTAGGTTTTTGATA 604
                |: | |:| || : ::|:|: :: ||: | | :| |: ::|
Db   692 ARGCNYTNAARTGGGARYTNTGYCNGTNACNGTNATHWSNTCGYTNAAYYTNNTYYTNC 751

Qy   605 AGGTCATAGAGAAGAGATACTTTTTAAGAG 634
                :::: ||:|| | |:| | ::: :|||
Db   752 ARGTNGAYGCNGTNAARGAYGTNCCNAARG 781

RESULT 6
US-09-222-851-18
; Sequence 18, Application US/09222851
; Patent No. 6165753
; GENERAL INFORMATION:
; APPLICANT: Coats, Steven R.
; APPLICANT: Bass, Michael B.
; APPLICANT: Robinson, Murray O.
; TITLE OF INVENTION: No. 6165753el Cyclin B Genes and Proteins
; FILE REFERENCE: A-524
; CURRENT APPLICATION NUMBER: US/09/222,851
; CURRENT FILING DATE: 1998-12-30
; EARLIER APPLICATION NUMBER: 09/092,770
; EARLIER FILING DATE: 1998-06-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 18
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Mouse
US-09-222-851-18

Query Match   4.9%;   Score 34.8; DB 3; Length 1212;
Best Local Similarity   29.3%;   Pred. No. 0.4;
Matches   79;   Conservative 40;   Mismatches 151;   Indels   0;   Gaps   0;

Qy   365 GAGTCTTGGTGACAAATGGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATACCAC 424
                | || || :   :   :| ||: || : | || ::
Db   512 GNTTYATGYTNACNCARAARGAYGTNAAYAARAAYATGYTNACARYTNARYTNGGNATHACNW 571

Qy   425 TGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTTTAA 484
                :: :   ::: |: |:: |||| :: |: || :|| ||:
Db   512 GNTTYATGYTNACNCARAARGAYGTNAAYAARAAYATGYTNACARYTNARYTNGGNATHACNW 571

Qy   425 TGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTTTAA 484
                :: :   ::: |: |:: |||| :: |: || :|| ||:
Db   572 SNYTNTTYATHGCNWSNAARYTNGARGARATHTAYGCNCNAARYTNCARGARYTNCARGCNT 631

```
Qy   485 CTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATTGGACTGCACAA 544
             |:||      |||   |:||:   :   :   |:||   |:|:   : :|
Db   632 AYGTNACNGAYGGNGNCNTGYWSNGARGTNGAYATHYTNAARATGGARYTNAAYATHYTNA 691

Qy   545 ATCCACGGTACTTAGCTGCAGGTTCTTCCAATGCTGTCAATGCTGTCAGTAGGTTTTTGATA 604
             |:|:      |||   :||      |:|      :||:
Db   692 ARGCNYTNAARTGGGARYTNTGYCCNGTNACNGTNATHWSNTGGYTNAAYTNTTYYTNC 751

Qy   605 AGGTCATAGAGAAGAAGAGATACTTTTTAAGAG 634
             |:||      |      ||:
Db   752 ARGTNGAYGCNGTNAARGAYGTNCCNAARG 781

RESULT 7
US-09-248-796A-779
; Sequence 779, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 779
; LENGTH: 1182
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-779

Query Match      4.8%;  Score 34.4;  DB 4;  Length 1182;
Best Local Similarity  48.5%;  Pred. No. 0.54;
Matches  95;  Conservative  0;  Mismatches 101;  Indels  0;  Gaps  0;

Qy   381 TGGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATACCACTGTTGGTAATAGGGAC 440
             |-||||      ||||      |||   |||
Db   582 TCGAGAACAAGATGAATCAATGATTAATTGATTAAACAAATTGAACAAATTGAAGCTGGACT 641

Qy   441 TAAACTGGACCAGATTCATGAAACAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCT 500
             |||||      ||      ||||      |||||      ||
Db   642 TAAAAATATTACCATGTCAGATATAAAAACGTCAACAATTACTTGATAAGAGAGATTCTTT 701

Qy   501 GGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGCGACAAATCCACGGTACTTAGC 560
             ||||||      |||||      ||:
Db   702 ATTACCGGATTATAGAATTGATTAATTTTATCGGATTTATCTAAACCAGGTTATCAACA 761

Qy   561 TGCAGGTTCTTCCAAT 576
             ||||||      |
Db   762 ATCAGGATTTTATGAT 777

RESULT 8
US-09-248-796A-6393
; Sequence 6393, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS 6393
```

```
; LENGTH: 810
; TYPE: DNA
; ORGANISM: Candida albicans
US-09-248-796A-6393

Query Match      4.8%;  Score 34.2;  DB 4;  Length 810;
Best Local Similarity  48.2%;  Pred. No. 0.5;
Matches  96;  Conservative  0;  Mismatches 103;  Indels  0;  Gaps  0;

Qy   373 GTGACAAATGGGATTATGATCAAGAACAGTTGCTGATAACCAAATACCACTGTTGGTA 432
             |||      ||||      |||      ||||||
Db   562 GTGGACAATACAGAACGGAGGAAGAAAATAATGATGATACAGACGAAATACCAAAATCCGAA 621

Qy   433 ATAGGGACTAAACTGGACCAGATTCATGAAACAAGCGCCATGAAGTTTAACTAGGACT 492
             |||      ||      ||||      |:||      ||||      |||||
Db   622 ACATTGAAACAAACAGGAGGAAGTTGGAATTGGACTGCGCAGTTAAAACTATCAAG 681

Qy   493 GCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATTCGGCGAAATGCCACAAATCCACGG 552
             |||      |||      |:||      |||||
Db   682 ACTATGGATGGAGAAATGAAAACTATTCGGCGAAATGTTACTGATTCAAATTTGGTAAA 741

Qy   553 TACTTAGTCGCAGGTTCTT 571
             ||||||      |
Db   742 GAATTGGGTGAAGGTTCCT 760

RESULT 9
US-09-270-767-13913
; Sequence 13913, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13913
; LENGTH: 735
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-09-270-767-13913

Query Match      4.8%;  Score 34;  DB 4;  Length 735;
Best Local Similarity  51.3%;  Pred. No. 0.55;
Matches  79;  Conservative  0;  Mismatches  75;  Indels  0;  Gaps  0;

Qy   558 AGCTGCAGGTTCTTCCAATGCTGTCAAGCTCAGTAGGTTTTTGATAGGTCATAGAGAA 617
             ||||||      ||      |||      |||||
Db   462 AGTTGCAGTCGCGGCGACATCGCCCCCTCTCCGTGGTTACGTCGGCAACTGCAAACGCAA 521

Qy   618 GAGATACTTTTTAAGAGAAGGTAATCAGATTCCAGGCTTCCTGATCGGAAAAGATTTGG 677
             |||||      |||||
Db   522 ATGCAAAGTGCCGCGGCAAGTTAATTTAATTCGGTTCTATTGTGATTGGAAAAGACATAG 581

Qy   678 GGCAGGAACATTAAAGAGACCTTCATTATGCTGA 711
             ||||      ||||||
Db   582 CGCAGCAGCAGGATGGATTCAAATTCTTAATGA 615

RESULT 10
US-09-134-000C-2635
; Sequence 2635, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
```

; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2635
; LENGTH: 2742
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-2635

Query Match   4.8%;   Score 33.8;   DB 4;   Length 2742;
Best Local Similarity   51.0%;   Pred. No. 1.5;
Matches   80; Conservative   0; Mismatches   77; Indels   0; Gaps   0;

Qy   515 ATCCAGAAGAAATTAATTGGACTGCACAAATCCACGGTACTTAGCTGCAGGTTCTTCCA 574
          ||||| ||| |||| |||| |||| ||| |||||||||| |||| |||||||||||||
Db   1346 ATCCAAAAGCCGTTAAGTTAAGAAACCTCATTGACCCCAGAGTTCGATGTCACTTACGATG 1405

Qy   575 ATGCTGTCAAGCTCAGTAGGTTTTTGATAAGGTCATAGAGAAGAGATACTTTTTAAGAG 634
          |||| | ||| |||| ||||||||||||| |||||||||||||||||||
Db   1406 GTGCCCATAATTATATGTTTTTTATGATGAAGTGAAAGAAATGACCACTACTATTCCAG 1465

Qy   635 AAGGTAATCAGATTCCAGGCGTTCCTGATCGGAAAAG 671
          |||| | ||||| |||||||||||||||
Db   1466 AAGTTACCTATAAGTTGGCTTGTTGTTGATGAAAAAGG 1502

RESULT 11
US-09-949-016-12690
; Sequence 12690, Application US/0994016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12690
; LENGTH: 45090
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-12690

Query Match   4.8%;   Score 33.8;   DB 4;   Length 45090;
Best Local Similarity   50.3%;   Pred. No. 9.6;
Matches   83; Conservative   0; Mismatches   82; Indels   0; Gaps   0;

Qy   6 GTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTT 65
          |||||||||| |||||||| | |||| ||||| || |||||| ||| |||||||||||
Db   2192 GACCTCCGCACGGGTATATGGGATGGAAGCGGGACCCTCGGGAGCAGGTAAGGTCCCCA 2251

Qy   66 AGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGACTGTGGGCTGCTC 125
          ||||||| ||||| ||||||||||| |||||| ||||||||||||||||| | |||||
Db   2252 GGAGGGGAACGGTGCATGCTCCAAGGACTGGGGGATCCCGCATGAAAAGCGTGGTTTCC 2311

Qy   126 AGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAGA 170
          ||||| |||||||||||||| ||||||
Db   2312 AAGTGATGGAAGCGCGTCCTGAGTGAGGAGAAGGGCTCTCCCACGA 2356

RESULT 12
US-09-949-016-16195
; Sequence 16195, Application US/0994016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16195
; LENGTH: 45091
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-16195

Query Match   4.8%;   Score 33.8;   DB 4;   Length 45091;
Best Local Similarity   50.3%;   Pred. No. 9.6;
Matches   83; Conservative   0; Mismatches   82; Indels   0; Gaps   0;

Qy   6 GTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTT 65
          |||||||||| |||||||| | |||| ||||| || |||||| ||| |||||||||||
Db   2192 GACCTCCGCACGGGTATATGGGATGGAAGCGGGACCCTCGGGAGCAGGTAAGGTCCCCA 2251

Qy   66 AGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGACTGTGGGCTGCTC 125
          ||||||| ||||| ||||||||||| |||||| ||||||||||||||||| | |||||
Db   2252 GGAGGGGAACGGTGCATGCTCCAAGGACTGGGGGATCCCGCATGAAAAGCGTGGTTTCC 2311

Qy   126 AGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAGA 170
          ||||| |||||||||||||| ||||||
Db   2312 AAGTGATGGAAGCGCGTCCTGAGTGAGGAGAAGGGCTCTCCCACGA 2356

RESULT 13
US-09-134-000C-2379
; Sequence 2379, Application US/09134000C
; Patent No. 6611156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2379
; LENGTH: 1008
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-2379

Query Match   4.7%;   Score 33.2;   DB 4;   Length 1008;
Best Local Similarity   57.8%;   Pred. No. 1.3;
Matches   59; Conservative   0; Mismatches   43; Indels   0; Gaps   0;

Qy   392 ATCAAGAACAGTTGCTGATAACCAAATACCACTGTTGGTAATAGGACTAAACTGGACC 451
          || |||| |||||| ||||| |||||||| ||| ||||||||||||||
Db   437 ATTTAGTCCGGTTGGGGAAGACCAAAAACCAATGTTGGAACAAGCCAAGAAATTGTGC 496

Qy   452 AGATTCATGAAACAAAGCGCCATGAAGTTTAACTAGGACTG 493
          ||| |||||||||| |||| |||||||||
Db   497 ATAGTTTTAATCAGACGTACGGTGAAGTTTTAGTTAGACCGG 538

RESULT 14
US-09-949-016-17513

```
Db  248  GKKRARSYKGSYKGGSYTYYKGGKKGGGTKGGKAMSKKGKGRRRAMTTTTYYKKKKWRSYWA 189
             :::  ::   :::||  |:::||:||:       |:       ::   ::   ::::::|
Qy  411  TAACCAAATACCACTGTTGGTAATAGGGACTAAACTGACCAGATTCATGAAACAAAGCG 470
             :|     :::      :::   ||   :::||   :::|:   :::  :::  ||
Db  188  ARGKTYKKWAAMMMMCMAWYMRSMMYYYKKWAAAWKSCCMARKKGGSSGRWYMWKRRG 129

Qy  471  CCATGAAGTTTAACTAGGACTGCTTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAA 530
             ||  |:|::| :::  ::: :: |:  :::  ||||   ||  |:|:|
Db  128  TCMRGRRRWYSRRRMCMWYCTKGSMAMMMSGKKRAAMCCCSKYYYTWMCMAAAAWWTWA 69
```

Search completed: June 30, 2005, 17:01:03
Job time : 139.331 secs

```
; Sequence 17513, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig et al.
;  TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
;  TITLE OF INVENTION:  WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
;  FILE REFERENCE: CL001307
;  CURRENT APPLICATION NUMBER: US/09/949,016
;  CURRENT FILING DATE: 2000-04-14
;  PRIOR APPLICATION NUMBER: 60/241,755
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/237,768
;  PRIOR FILING DATE: 2000-10-03
;  PRIOR APPLICATION NUMBER: 60/231,498
;  PRIOR FILING DATE: 2000-09-08
;  NUMBER OF SEQ ID NOS: 207012
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17513
;  LENGTH: 121049
;  TYPE: DNA
;  ORGANISM: Human
US-09-949-016-17513

Query Match            4.6%;  Score 33;  DB 4;  Length 121049;
Best Local Similarity   51.0%;  Pred. No. 35;
Matches  78;  Conservative  0;  Mismatches  75;  Indels  0;  Gaps  0;

Qy     384  GGATTATGATCAAGAACACAGTTGCTGATAACCAAATACCACTGTTGGTAATAGGGACTAA 443
            | ||  ||||     || |||||   ||   || ::   ||||| |||||||||||||||
Db  108039  GTATAATGATGTTAATCTTTTATTTCAGAATCTTCAATCATTGTTTCTTGTAGGGAATAA 108098

Qy     444  ACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTTTAACTAGGACTGCTTTCCTGGC 503
            | |||  |||||||  ||||| ||||  ||||||||||| ||   |||| |||||||||
Db  108099  AATAAAACTATTTGTATAGCTTTACCCCCATAAAAATTTACCCTTTAACTGTTTCTGCCTGT 108158

Qy     504  TGAGGATTTCAATCCAGAAGAAATTAATTTGGA 536
            ||||||||  | |||||||||| ||||| ||||
Db  108159  AGGACAGGGTAACCTTCAGGAGAATAATTGTGA 108191
```

RESULT 15
US-09-621-976-18033/c
```
; Sequence 18033, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;  APPLICANT: Dumas Milne Edwards, J.B.
;  APPLICANT: Jobert, S.
;  APPLICANT: Giordano, J.Y.
;  TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;  FILE REFERENCE: GENSET.054PR2
;  CURRENT APPLICATION NUMBER: US/09/621,976
;  CURRENT FILING DATE: 2000-07-21
;  NUMBER OF SEQ ID NOS: 19335
;  SOFTWARE: Patent.pm
; SEQ ID NO 18033
;  LENGTH: 474
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: misc_feature
;  LOCATION: 16
;  OTHER INFORMATION: n=a, g, c or t
US-09-621-976-18033

Query Match            4.6%;  Score 32.8;  DB 4;  Length 474;
Best Local Similarity   12.9%;  Pred. No. 1.1;
Matches  31;  Conservative  106;  Mismatches  103;  Indels  0;  Gaps  0;

Qy     291  AACAAATAAGAGAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGA 350
            ||  ||:: :||::|  :::  || : ::: :::::::::|  :: |:: |::  ::::
Db     308  AARRWTWARRAAMYTWWTYMWKGSCMWKSMWMRKWRKYMWKCCYWWRKYCCMRKKRRRR 249

Qy     351  TTTGGTGCCAACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTTGCTGA 410
```

```
                    GenCore version 5.1.6
              Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:        June 30, 2005, 16:53:46 ; Search time 492.231 Seconds
                                         (without alignments)
                                         9059.822 Million cell updates/sec

Title:         US-09-945-173-3
Perfect score: 711
Sequence:      1 atggcgtccctggatcggt.........agagccttcattatgactga 711

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      613374 seqs, 3136092125 residues

Total number of hits satisfying chosen parameters:        12626748

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications NA:*
              1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
              2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
              3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
              4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
              5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
              6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
              7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
              8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
              9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
             10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
             11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
             12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
             13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
             14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
             15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
             16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq:*
             17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq:*
             18: /cgn2_6/ptodata/2/pubpna/US10F_PUBCOMB.seq:*
             19: /cgn2_6/ptodata/2/pubpna/US10G_PUBCOMB.seq:*
             20: /cgn2_6/ptodata/2/pubpna/US10H_PUBCOMB.seq:*
             21: /cgn2_6/ptodata/2/pubpna/US10I_PUBCOMB.seq:*
             22: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
             23: /cgn2_6/ptodata/2/pubpna/US11A_PUBCOMB.seq:*
             24: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq:*
             25: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
             26: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

```
Result         %      Query
No.    Score  Match  Length  DB  Length  ID                      Description
-----  -----  -----  ------  --  ------  --------------------    -----------
1      711    100.0    711    9    711   US-09-945-173-3         Sequence 3, Appli
2      711    100.0    711   17    735   US-10-210-130-125       Sequence 125, App
3      711    100.0    711   13   1417   US-10-044-090-467       Sequence 467, App
4      711    100.0    711    9   1694   US-09-945-173-1         Sequence 1, Appli
5      711    100.0    711   19   2398   US-10-479-284-34        Sequence 34, Appl
6      707.8   99.5    739   17   2398   US-10-210-130-127       Sequence 127, App
7      617.4   86.8    959    9    959   US-09-764-868-321       Sequence 321, App
```

```
                                                               8   444  62.4      504  10   US-09-918-995-995-23516         Sequence 23516, A
                                                               9   54.6  7.7     1626  20   US-10-425-115-2369              Sequence 2369, Ap
                                                              10   51.4  7.2      664  19   US-10-767-701-25416            Sequence 25416, A
                                                              11   50.2  7.1     1643  18   US-10-425-114-26886            Sequence 26886, A
                                                              12   50.2  7.1     1643  20   US-10-425-115-75255            Sequence 75255, A
                                                              13   45.2  6.4      761  19   US-10-437-963-650             Sequence .650, App
                                                              14   44.6  6.3     1584  19   US-10-437-963-25042           Sequence 25042, A
                                                              15   38    5.3      610  18   US-10-424-599-109350          Sequence 109350,
                                                              16   37.4  5.3     2693  19   US-10-437-963-64985           Sequence 64985, A
                                                              17   36.4  5.1     1510  18   US-10-424-599-109349          Sequence 109349,
                                                              18   36.2  5.1      648  21   US-10-487-901-5702            Sequence 5702, Ap
                                                              19   36.2  5.1     1161  19   US-10-437-963-64599           Sequence 64599, A
                                                              20   36.2  5.1     3836  19   US-10-437-963-20359           Sequence 20359, A
                                                              21   35.6  5.0      775  14   US-10-146-731-120             Sequence 120, App
                                                              22   35.6  5.0      775  15   US-10-140-472-120             Sequence 120, App
                                                              23   35.6  5.0      775  15   US-10-141-761-120             Sequence 120, App
                                                              24   35.6  5.0      775  15   US-10-142-85-120              Sequence 120, App
                                                              25   35.6  5.0      775  16   US-10-142-88-120              Sequence 120, App
                                                              26   35.6  5.0      775  16   US-10-158-790-120             Sequence 120, App
                                                              27   35.6  5.0      775  16   US-10-137-871-120             Sequence 120, App
                                                              28   35.6  5.0      775  17   US-10-140-922-120             Sequence 120, App
                                                              29   35.6  5.0      775  17   US-10-141-756-120             Sequence 120, App
                                                              30   35.6  5.0      775  17   US-10-141-759-120             Sequence 120, App
                                                              31   35.6  5.0      775  17   US-10-140-805-120             Sequence 120, App
                                                              32   35.6  5.0      775  17   US-10-140-864-120             Sequence 120, App
                                                              33   35.6  5.0      775  18   US-10-142-426-120             Sequence 120, App
                                                              34   35.4  5.0   165097  22   US-10-737-08-77                Sequence 77, Appl
                                                              35   35.4  5.0   165097  22   US-10-765-790-77               Sequence 77, Appl
                                                              36   34.8  4.9      468  14   US-10-123-155-90              Sequence 90, Appl
                                                              37   34.8  4.9      468  15   US-10-146-731-90             Sequence 90, Appl
                                                              38   34.8  4.9      468  15   US-10-140-472-90             Sequence 90, Appl
                                                              39   34.8  4.9      468  15   US-10-141-761-90             Sequence 90, Appl
                                                              40   34.8  4.9      468  16   US-10-142-85-90             Sequence 90, Appl
                                                              41   34.8  4.9      468  16   US-10-158-790-90             Sequence 90, Appl
                                                              42   34.8  4.9      468  16   US-10-063-685-150            Sequence 150, App
                                                              43   34.8  4.9      468  17   US-10-137-87-90             Sequence 90, Appl
                                                              44   34.8  4.9      468  17   US-10-140-923-90            Sequence 90, Appl
                                                              45   34.8  4.9      468  17   US-10-141-756-90            Sequence 90, Appl
```

ALIGNMENTS

```
RESULT 1
US-09-945-173-3
; Sequence 3, Application US/09945173
; Patent No. US20022012756BA1
; GENERAL INFORMATION:
;    APPLICANT: Millennium Pharmaceuticals, Inc.
;    TITLE OF INVENTION: Meyers, Rachel
;    TITLE OF INVENTION: 47324, A NOVEL HUMAN G-PROTEIN AND USES
;    TITLE OF INVENTION:  THEREFOR
;    FILE REFERENCE: 38155-20035.00
;    CURRENT APPLICATION NUMBER: US/09/945,173
;    CURRENT FILING DATE:  2001-08-31
;    PRIOR APPLICATION NUMBER: US 60/229,293
;    PRIOR FILING DATE: 2000-09-01
;    NUMBER OF SEQ ID NOS: 11
;    SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;    LENGTH: 711
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-09-945-173-3

  Query Match             100.0%;  Score 711; DB 9; Length 711;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-213;
  Matches  711  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1 ATGGCGTCCCTGGATCGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCT 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 ATGGCGTCCCTGGATCGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCT 60
```

```
Qy   61  TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGGC 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGGC 120

Qy  121  TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAGAAGACCTAC 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAGAAGACCTAC 180

Qy  181  TACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGAAAAAGCACA 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  TACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGAAAAAGCACA 240

Qy  241  AGAGCAGTATTCTACAACTCCGTAAATGTATTATTTCGTACACGACTTAACAAATAAG 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  AGAGCAGTATTCTACAACTCCGTAAATGTATTATTTCGTACACGACTTAACAACAAATAAG 300

Qy  301  AAGTCCTCCCAAAACTTGGCTCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCA 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  AAGTCCTCCCAAAACTTGGCTCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCA 360

Qy  361  ACTGGAGTCTTGGTGACAAATGGGGATTATGATGATCAAGAACAGTTTGCTGATAACCAAATA 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  ACTGGAGTCTTGGTGACAAATGGGGATTATGATGATCAAGAACAGTTTGCTGATAACCAAATA 420

Qy  421  CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCAATGAAACAAAGGCGCCATGAAGTT 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421  CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCAATGAAACAAAGGCGCCATGAAGTT 480

Qy  481  TTAACTAGGACTGCTTCCTGGCTTCAATCCAGAAGAAATTAATTTGGACTGC 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481  TTAACTAGGACTGCTTCCTGGCTTCAATCCAGAAGAAATTAATTTGGACTGC 540

Qy  541  ACAAATCCACGGTACTTAGCTGCGAGGTTCTTCCCAATGCTGTCAAGCTCAGTAGGTTTTT 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541  ACAAATCCACGGTACTTAGCTGCGAGGTTCTTCCCAATGCTGTCAAGCTCAGTAGGTTTTT 600

Qy  601  GATAAGGTCATAGAGAAGAAGATACTTTTTAAGAGAGAGGTAATCAGATTCCAGGCTTCCT 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601  GATAAGGTCATAGAGAAGAAGATACTTTTTAAGAGAGAGGTAATCAGATTCCAGGCTTCCT 660

Qy  661  GATCGGAAAAGATTTGGGGCCAGGAACATTAAAGAGCCTTCATTATGACTGA 711
         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661  GATCGGAAAAGATTTGGGGCCAGGAACATTAAAGAGCCTTCATTATGACTGA 711
```

```
RESULT 2
US-10-210-130-125
; Sequence 125, Application US/10210130
; Publication No. US2004014053A1
GENERAL INFORMATION:
APPLICANT: Zerhusen, Bryan D.
APPLICANT: Patturajan, Meera
APPLICANT: Kekuda, Ramesh
APPLICANT: Miller, Charles E.
APPLICANT: Rieger, Daniel K.
APPLICANT: Pena, Carol E.A.
APPLICANT: Shimkets, Richard A.
APPLICANT: Li, Li
APPLICANT: Berghs, Constance
APPLICANT: Zhong, Mei
APPLICANT: Casman, Stacie J.
APPLICANT: Voss, Edward -Z.
APPLICANT: Boldog, Ferenc L.
APPLICANT: Padigaru, Muralidhara
APPLICANT: Smithson, Glennda
APPLICANT: Ji, Weizhen
APPLICANT: Gorman, Linda
APPLICANT: Vernet, Corine A.M.
APPLICANT: Leite, Mario W.
APPLICANT: Guo, Xiaojia Sasha
APPLICANT: Anderson, David W.
APPLICANT: Spytek, Kimberly A.
APPLICANT: Gerlach, Valerie
```

```
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Khramtsov, Nikolai V.
; APPLICANT: Ort, Tatiana
; APPLICANT: Ellerman, Karen
; APPLICANT: Rastelli, Luca
; APPLICANT: Agee, Michele L.
; APPLICANT: Chaudhuri, Amitabha
; APPLICANT: Chant, John S.
; APPLICANT: DiPippo, Vincent A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Eisen, Andrew J.
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Giot, Loic
; APPLICANT: Ooi, Chean Eng
; APPLICANT: Rothenberg, Mark E.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Hjalt, Tord
; APPLICANT: Taupier, Raymond J., Jr.
; APPLICANT: Catterton, Elina
; APPLICANT: Shenoy, Suresh G.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-416C (Cura-716 SMT)
; CURRENT APPLICATION NUMBER: US/10/210,130
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: 60/309,501
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 60/316,508
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: 60/354,655
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: 60/310,291
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 60/383,887
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: 60/310,951
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/323,936
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: 60/381,039
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/311,292
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/311,979
; PRIOR FILING DATE: 2001-08-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 369
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 125
; LENGTH: 735
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10)..(717)
US-10-210-130-125

Query Match        100.0%;  Score 711;  DB 17;  Length 735;
Best Local Similarity  100.0%;  Pred. No. 1.5e-213;
Matches  711;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1  ATGGGCTCCCTGGATCGGGTCGAAGTGCTGGGAGACTCAGTGTGTTGGGAAATCT  60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10  ATGGGCTCCCCTGGATCGGGTCGAAGTGCTGGGAGACTCAGTGTGTTGGGAAATCT  69

Qy   61  TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGGC 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70  TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGGC 129

Qy  121  TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAGAAGACCTAC 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  130  TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAGAAGACCTAC 189
```

Qy   181 TACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGCTGAAAAGCACA 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   183 TACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGCTGAAAAGCACA 242

Qy   241 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGACTTAACAAATAAG 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   243 AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGACTTAACAAATAAG 302

Qy   301 AAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTGGTGCCA 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   303 AAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTGGTGCCA 362

Qy   361 ACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTGCTGATAACCAAATA 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   363 ACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTGCTGATAACCAAATA 422

Qy   421 CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTT 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   423 CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTT 482

Qy   481 TTAACTAGGACTGCTTTCCTGCTGAGGATTTCAATCCAGAGAAATTAATTTGGACTGC 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   483 TTAACTAGGACTGCTTTCCTGCTGAGGATTTCAATCCAGAGAAATTAATTTGGACTGC 542

Qy   541 ACAAATCCACGTACTTAGCTGCAGGTTCTTCCAATGCTGTGTCAAGCTCAGTAGGTTTTT 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   543 ACAAATCCACGTACTTAGCTGCAGGTTCTTCCAATGCTGTGTCAAGCTCAGTAGGTTTTT 602

Qy   601 GATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAATCAGATTCCAGGCTTTCCT 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   603 GATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAATCAGATTCCAGGCTTTCCT 662

Qy   661 GATCGGAAAAGATTTGGGCAGGAACATTAAAGAGCCTTCATTATGACTGA 711
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   663 GATCGGAAAAGATTTGGGCAGGAACATTAAAGAGCCTTCATTATGACTGA 713

RESULT 4
US-09-945-173-1
; Sequence 1, Application US/09945173
; Patent No. US2002012756BA1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 47324, A NOVEL HUMAN G-PROTEIN AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 38155-20035.00
; CURRENT APPLICATION NUMBER: US/09/945,173
; CURRENT FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/229,293
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1694
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19)...(729)
US-09-945-173-1

Query Match   100.0%; Score 711; DB 9; Length 1694;
Best Local Similarity 100.0%; Pred. No. 2.4e-213;
Matches 711; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 ATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGTGTTGGGAAATCT 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   19 ATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGTGTTGGGAAATCT 78

Qy   61 TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGCGAACTGTGGGC 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db ...

RESULT 3
US-10-044-090-467
; Sequence 467, Application US/10044090
; Publication No. US2002013708A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 467
; LENGTH: 1417
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US2002013708A1 1307098.1
US-10-044-090-467

Query Match   100.0%; Score 711; DB 13; Length 1417;
Best Local Similarity 100.0%; Pred. No. 2.2e-213;
Matches 711; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 ATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGTGTTGGGAAATCT 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3 ATGGCGTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGTGTTGGGAAATCT 62

Qy   61 TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGCGAACTGTGGGC 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63 TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGCGAACTGTGGGC 122

Qy   121 TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAGAGAAGACCTAC 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   123 TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAGAGAAGACCTAC 182

```
Db    79  TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGGC  138

Qy   121  TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAAGAACCCCAGAAGAGAGAAGACCTAC  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   139  TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAACCCCAGAAGAGAGAAGACCTAC  198

Qy   181  TACATAGAATTATGGGATGTTGGAGGCTCTGTGGCAGTGCCAGCAGCGGTGAAAAGCACA  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   199  TACATAGAATTATGGGATGTTGGAGGCTCTGTGGCAGTGCCAGCAGCGGTGAAAAGCACA  258

Qy   241  AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGACTTAACAAATAAG  300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   259  AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGACTTAACAAATAAG  318

Qy   301  AAGTCCTCCCAAAACTTGGCTGGTCGTTGGTCTCAACAGTTGCTGATAACCAAATA  360
          ||||||||||||||||||||||||||||||||||||||||||||
Db   319  AAGTCCTCCCAAAACTTGGCTGGTCGTTGGTCTCAACAGTTGCTGATAACCAAATA  378

Qy   361  ACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATA  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   379  ACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATA  438

Qy   421  CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTT  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   439  CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTT  498

Qy   481  TTAACTAGGACTGCTTCCTGCTGAGGATTCAATCCAGAAGAAAATTAATTTGGACTGC  540
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   499  TTAACTAGGACTGCTTCCTGCTGAGGATTCAATCCAGAAGAAAATTAATTTGGACTGC  558

Qy   541  ACAAATCCACGTACTTAGCTGCAGTTCGTCAGCTCGTCAGTAGGTTTTTT  600
          |||||||||||||||||||||||||||||||||||||||||
Db   559  ACAAATCCACGTACTTAGCTGCAGTTCGTCAGCTCGTCAGTAGGTTTTTT  618

Qy   601  GATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAATCAGATTCCAGGCTTTCCT  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   619  GATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAATCAGATTCCAGGCTTTCCT  678

Qy   661  GATCGGAAAAGATTTGGGGCAGGAGAACATTAAAGAGAGCCTTCATTATGACTGA  711
          ||||||||||||||||||||||||||||||||||||||||||||
Db   679  GATCGGAAAAGATTTGGGGCAGGAGAACATTAAAGAGAGCCTTCATTATGACTGA  729
```

```
; SEQ ID NO 34
;     LENGTH: 2398
;     TYPE: DNA
;     ORGANISM: Homo sapiens
;     FEATURE:
;     NAME/KEY: misc feature
;     OTHER INFORMATION: Incyte ID No: 747515CB1
US-10-479-284-34

Query Match        100.0%;  Score 711;  DB 19;  Length 2398;
Best Local Similarity  100.0%;  Pred. No. 2.9e-213;
Matches  711;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;
```

```
Qy     1  ATGGGCTCCCTGCCGATCGGGGTGAAGGTACTGGTGTTGGGAGACTCAGTGTGTTGGGAAATCT  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3  ATGGGCGTCCCTGCCGATCGGGGTGAAGGTACTGGTGTTGGGAGACTCAGTGTGTTGGGAAATCT  62

Qy    61  TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGGC  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    63  TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGGC  122

Qy   121  TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAAGAACCCCAGAAGAGAAGAACCTAC  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   123  TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAAGAACCCCAGAAGAGAAGAACCTAC  182

Qy   181  TACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGAAAAGCACA  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   183  TACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGAAAAGCACA  242

Qy   241  AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGACTTAACAAATAAG  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   243  AGAGCAGTATTCTACAACTCCGTAAATGGTATTATTTCGTACACGACTTAACAAATAAG  302

Qy   301  AAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCA  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   303  AAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCA  362

Qy   361  ACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATA  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   363  ACTGGAGTCTTGGTGACAAATGGGGATTATGATCAAGAACAGTTTGCTGATAACCAAATA  422

Qy   421  CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTT  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   423  CCACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAGCGCCATGAAGTT  482

Qy   481  TTAACTAGGACTGCTTCCTGGCTGGATTTCAATCCAGAGAAGAAATTAATTTGGACTGC  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   483  TTAACTAGGACTGCTTCCTGGCTGGATTTCAATCCAGAGAAGAAATTAATTTGGACTGC  542

Qy   541  ACAAATCCACGTACTTAGCTGCAGTTCGTCAAGTGCTGTCAAGCTCAGTAGGTTTTTT  600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   543  ACAAATCCACGTACTTAGCTGCAGTTCGTCAAGTGCTGTCAAGCTCAGTAGGTTTTTT  602

Qy   601  GATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAATCAGATTCCAGGCTTTCCT  660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   603  GATAAGGTCATAGAGAAGAGATACTTTTTAAGAGAAGGTAATCAGATTCCAGGCTTTCCT  662
```

```
Qy  661 GATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGAGCCTTCATTATGACTGA 711
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  663 GATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGAGCCTTCATTATGACTGA 713
```

RESULT 6
US-10-210-130-127
; Sequence 127, Application US/10210130
; Publication No. US20040014053A1
; GENERAL INFORMATION:
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Patturajan, Meera
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Miller, Charles E.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Pena, Carol E.A.
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Li, Li
; APPLICANT: Berghs, Constance
; APPLICANT: Zhong, Mei
; APPLICANT: Casman, Stacie J.
; APPLICANT: Voss, Edward Z.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Smithson, Glennda
; APPLICANT: Ji, Weizhen
; APPLICANT: Gorman, Linda
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Leite, Mario W.
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Anderson, David W.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Khramtsov, Nikolai V.
; APPLICANT: Ort, Tatiana
; APPLICANT: Ellerman, Karen
; APPLICANT: Rastelli, Luca
; APPLICANT: Agee, Michele L.
; APPLICANT: Chaudhuri, Amitabha
; APPLICANT: Chant, John S.
; APPLICANT: DiPippo, Vincent A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Eisen, Andrew J.
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Giot, Loic
; APPLICANT: Ooi, Chean Eng
; APPLICANT: Rothenberg, Mark E.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Hjalt, Tord
; APPLICANT: Liu, Xiaohong
; APPLICANT: Taupier, Raymond J., Jr.
; APPLICANT: Catterton, Elina
; APPLICANT: Shenoy, Suresh G.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-416C (Cura-716 SMT)
; CURRENT APPLICATION NUMBER: US/10/210,130
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: 60/309,501
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: 60/316,508
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: 60/354,655
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: 60/310,291
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 60/383,887
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: 60/310,951
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/323,936
; PRIOR FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: 60/381,039
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/311,292
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/311,979
; PRIOR FILING DATE: 2001-08-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 369
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 127
;   LENGTH: 739
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (10)...(717)
US-10-210-130-127

Query Match   99.5%; Score 707.8; DB 17; Length 739;
Best Local Similarity 99.7%; Pred. No. 1.6e-212;
Matches 709; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1 ATGGCGTCCCTGGATCGGCGGTGAAGGTACTGGTGTGGAGGACTCAGGTGTTGGGAAATCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10 ATCGCGTCCCTGGATCGGCGGTGAAGGTACTGGTGTGGAGGACTCAGGTGTTGGGAAATCT 69

Qy   61 TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGGC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70 TCGTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGGC 129

Qy 121 TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAAGAACAACCCCAGAAGAGAGAACGACCTAC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 130 TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGAAGAACAACCCCAGAAGAGAGAACGACCTAC 189

Qy 181 TACATAGAATTATGGGATGTGGAGGCTCGTGTGGGCAGTGCCAGCAGCGTGAAAAGCACA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 190 TACATAGAATTATGGGATGTGGAGGCTCGTGTGGGCAGTGCCAGCAGCGTGAAAAGCACA 249

Qy 241 AGAGCAGTATTCTACAACTCCGTAAATGTATTATTTCGTACACGACTTAACAAATAAG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 250 AGAGCAGTATTCTACAACTCCGTAAATGTATTATTTCGTACACGACTTAACAAATAAG 309

Qy 301 AAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTGGTGCCA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 310 AAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTGGTGCCA 369

Qy 361 ACTGGAGTCTTGGTGACAAATGGGGATTAATGATCAAGAACAGTTTGCTGATAACCAAATA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 370 ACTGGAGTCTTGGTGACAAATGGGGATTAATGATCAAGAACAGTTTGCTGATAACCAAATA 429

Qy 421 CCACTGTTGGTAATAGGGACTAAACTGACCAGATTCATGAAACAAACAAAGCGCCATGAAGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 430 CCACTGTTGGTAATAGGGACTAAACTGACCAGATTCATGAAACAAACAAAGCGCCATGAAGT 489

Qy 481 TTAACTAGGACTGCTTTCCTGGCCGAGTTCTTCAAATGCTGTCAAGCTCAGTAGGTTTTT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 490 TTAACTAGGACTGCTTTCCTGGCCGAGTTCTTCAAATGCTGTCAAGCTCAGTAGGTTTTT 549

Qy 541 ACAAATCCACGTACTTAGCTGCAGTTCTCCCAAATGCTGTCAAGCTCAGTAGGTTTTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 550 ACAAATCCACGTACTTAGCTGCAGTTCTCCCAAATGCTGTCAAGCTCAGTAGGTTTTT 609

Qy 601 GATAAGGTCATAGAGAGAGATACTTTTTAAGAGAAGTAATCAGATTCCAGGCTTTCCT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 610 GATAAGGTCATAGAGAGAGATACTTTTTAAGAGAAGTAATCAGATTCCAGGCTTTCCT 669

Qy 661 GATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGAGCCTTCATTATGACTGA 711
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 670 GATCGGAAAAGATTTGGGGCAGGAACATTAAAGAGAGCCTTCATTATGACTGA 720
```

RESULT 7
US-09-764-868-321

```
;   Sequence 321, Application US/09764868
;   Patent No. US20020168711A1
;   GENERAL INFORMATION:
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PTZ32
;   CURRENT APPLICATION NUMBER: US/09/764,868
;   CURRENT FILING DATE: 2001-01-17
;   Prior application data removed - refer to PALM or file wrapper
;   NUMBER OF SEQ ID NOS: 1510
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 321
;   LENGTH: 959
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (652)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: SITE
;   LOCATION: (705)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-321

Query Match          86.8%;    Score 617.4; DB 9;    Length 959;
Best Local Similarity 96.4%;   Pred. No. 6.6e-184;   Indels   6;   Gaps   4;
Matches  673;   Conservative   1;   Mismatches  18;
```

```
Qy      1 ATGGCGTCCCTGGATCGGGTGAAGCTGGTGTTGGGAGACTCAGGTGTTGGGAAATCT   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12 ATGGCGGGTCCCTGGATCGGGTGAAGCTGGTGTTGGGAGACTCAGGTGTTGGGAAATCT   71

Qy     61 TCGTTAGTCCATCTCCTATGCCAAATCAGTGCTGGGAAATCATCATGGACTGTGGGC   120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     72 TCGTTAGTCCATCTCCTATGCCAAATCAGTGCTGGGAAATCATCATGGACTGTGGGC   131

Qy    121 TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAAGACCTAC   180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    132 TGCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAAGACCTAC   191

Qy    181 TACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGAAAAGCACA   240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    192 TACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGAAAAGCACA   251

Qy    241 AGAGCAGTATTCTACAACTCCGTAAATGTTATTATTTCGTACACGACTTAACAAATAAG   300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    252 AGAGCAGTATTCTCACAACTCCGTAAATGTTATTATTTCGTACACGACTTAACAAATAAG   311

Qy    301 AAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGTGCCA   360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    312 AAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGTGCCA   371

Qy    361 ACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTGCTGATAACCAAATA   420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    372 ACTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTGCTGATAACCAAATA   431

Qy    421 CCACTGTTGGTAATAGGACTAAACTGGACCAGATTCAATCAGAAGAAATTAATTTGGACTGC   480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    432 CCACTGTTGGTAATAGGACTAAACTGGACCAGATTCAATCAGAAGAAATTAATTTGGACTGC   491

Qy    481 TTAACTAGGACTGCTTCCTGGCTAGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGC   540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    492 TTAACTAGGACTGCTTCCTGGCTGAGGATTTCAATCCAGAAGAAATTAATTTGGACTGC   551

Qy    541 ACAAATCCACGTACTTAGCTGCGAGGTTCTCCAATGCTGTCAAGCTCAGTAGGTTTTTT   600
          |||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db    552 ACAAATCCACGTACTTAGCTGCGAGGTTCTCCAATGCTGTCAA-CTCARTAGGTTTTTT   610

Qy    601 GATAAGTCATAGAGAAGAGAGATACTTTTTAAGAGAAGTAATCAGATTCCAGGCTTCCT   660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    611 GAT-AGGGCATAGAGAAGAGAGATAC-TTTTAAGAGAAGTAATCNGATCCGAGGCTTCTTG   668

Qy    661 GATCGGAAAAGATTTGGGCAGGAACATTAAAGAGCCT   698
          |||||||||||||||||||||||||||||
Db    669 ACTCG---AAAAGATTGGGCCGGACCATTAAAGGCGCTT   703
```

```
RESULT 8
US-09-918-995-23516
;   Sequence 23516, Application US/09918995
;   Publication No. US20030073623A1
;   GENERAL INFORMATION:
;   APPLICANT: Hyseq, Inc.
;   TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
;   TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
;   FILE REFERENCE: 20411-756
;   CURRENT APPLICATION NUMBER: US/09/918,995
;   CURRENT FILING DATE: 2001-07-30
;   PRIOR APPLICATION NUMBER: US/09/235,076
;   PRIOR FILING DATE: 1999-01-20
;   NUMBER OF SEQ ID NOS: 38054
;   SOFTWARE: FastSEQ for Windows Version 3.0
;   SEQ ID NO 23516
;   LENGTH: 504
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(504)
;   OTHER INFORMATION: n = A,T,C or G
US-09-918-995-23516

Query Match           62.4%;    Score 444;    DB 10;    Length 504;
Best Local Similarity 96.4%;    Pred. No. 2.9e-129;
Matches  450;   Conservative   0;   Mismatches   17;   Indels   0;   Gaps   0;
```

```
Qy      2 TGGGCTCCCTGGATCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTT   61
          ||||||||||||    ||||||||   ||||||||||||||||||||||||||||||||||
Db     37 TTGGNANNNNGGCTCCGNNGAAGGGACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTT   96

Qy     62 CGTTAGTCCATCTCCTATGCCAAATCAAGTGCTGGGAAATCATCATGGACTGTGGGCT   121
          ||||||||||||||||||   |||||||||||||||||||||||||||||||||||||||
Db     97 CGTTAGTTCATCTCCTATGCCTAAATCAAGTGCTGGGAAATCATCATGGACTGTGGGCT   156

Qy    122 GCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAAGAGACTACT   181
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    157 GCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAAGAGACTACT   216

Qy    182 ACATAGAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGAAAAGCACAA   241
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    217 ACATATAATTATGGGATGTTGGAGGCTCTGTGGGCAGTGCCAGCAGCGTGAAAAGCACAA   276

Qy    242 GAGCAGTATTCTACAACTCCGTAAATGTATTATTTCGTACACGACTTAACAGACTTAAGA   301
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    277 GAGCAGTATTCTACAACTCCGTAAATGTATTATTTCGTACACGACTTAACAGACTTAAGA   336

Qy    302 AGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCAA   361
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    337 AGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCTCTCAACAGGGATTTGGTGCCAA   396

Qy    362 CTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTGCTGATAACCAAATAC   421
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    397 CTGGAGTCTTGGTGACAAATGGGATTATGATCAAGAACAGTTGCTGATAACCAAATAC   456

Qy    422 CACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAG   468
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    457 CACTGTTGGTAATAGGGACTAAACTGGACCAGATTCATGAAACAAAG   503
```

```
RESULT 9
US-10-425-115-2369
;   Sequence 2369, Application US/10425115
;   Publication No. US20040214272A1
;   GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
```

; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 2369
; LENGTH: 1626
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_102161C.1
US-10-425-115-2369

Query Match   7.7%; Score 54.6; DB 20; Length 1626;
Best Local Similarity   60.4%; Pred. No. 5.8e-06;
Matches   90; Conservative   0; Mismatches 59; Indels   0; Gaps   0;

```
Qy      3 GGCGTCCCTGGATCGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTC 62
            |||||||| ||||  || ||||||||||||||||||||||||||||||||||||||||
Db    342 GCCGCCCTGCGGCCAGGTTCGTGTACTCGTCGTTGGCGACTCAGGTGTTGGCAAATCTTC 401

Qy     63 GTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGGCTG 122
            || |||||||||||| |||  || |||||||||||||| |||||||||||||||||||
Db    402 ATTGGTGCATCTCTTACTGAAAGATTCTGCAGTGGCTGGCCAGCCAAACAATTGGGTG 461

Qy    123 CTCAGTGGATGTCAGAGTTCATGATTACA 151
            ||||||||||| ||||||||||| |||||
Db    462 TGCAGTTGGTGTTAAACATATTACTTACA 490
```

RESULT 10
US-10-767-701-25416
; Sequence 25416, Application US/10767701
; Publication No. US2004017268441
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
SEQ ID NO 25416
; LENGTH: 664
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 30951803
US-10-767-701-25416

Query Match   7.2%; Score 51.4; DB 19; Length 664;
Best Local Similarity   59.1%; Pred. No. 3.7e-05;
Matches   88; Conservative   0; Mismatches 61; Indels   0; Gaps   0;

```
Qy      3 GGCGTCCCTGGATCGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTC 62
            |||||||| ||||  |||: |||||||||||||||||||||||||||||||||||||
Db    398 GCCGCCCTGCGGCCAGGTTCGTGTGCTCGTCGTTGGCGACTCAGGGGTTGGTAAATCTTC 457

Qy     63 GTTAGTCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGGACTGTGGGCTG 122
            || |||||||||||| |||  ||  ||||||||||||| |||||||||||||||||||
Db    458 ATTGGTGCATCTCTTATTGAAAGGTTCTGCAGTGGCTCGACTGGCTGGCCAGCCAAACAATTGGGTG 517

Qy    123 CTCAGTGGATGTCAGAGTTCATGATTACA 151
            ||||||||||| ||||||||||| |||||
Db    518 TGCAGTTGGTGTTAAACATATCACTTACA 546
```

RESULT 11

US-10-425-114-26886
; Sequence 26886, Application US/10425114
; Publication No. US20040034880A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven B
; APPLICANT: Tabaška, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 26886
; LENGTH: 1643
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4666-008-F5_FLI
US-10-425-114-26886

Query Match   7.1%; Score 50.2; DB 18; Length 1643;
Best Local Similarity   59.4%; Pred. No. 0.00014;
Matches   85; Conservative   0; Mismatches 58; Indels   0; Gaps   0;

```
Qy      8 CCCTGGATCGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTTAG 67
            |||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    325 CCTTCGGCCAAGTGCGCGTCCTCATCGTCGTGATTCAGGTGTTGGGAAATCTTCATTGG 384

Qy     68 TCCATCTCCTATGCCAAAATCAAGTGCTGGCCATTGCTCGACCATCTCAAACACTAGGATGACACTG 127
            | ||||| |||||| |||||||||||||  || |||||| ||||| ||
Db    385 TGCATCTCATTTGAAAAGTTCTGCCATTGCTCGACCATCTCAAACACTAGGATGACACTG 444

Qy    128 TGGATGTCAGAGTTCATGATTAC 150
            |||  ||||||||||| ||||
Db    445 TGGGCATTAAACATGTTACTTAC 467
```

RESULT 12
US-10-425-115-75255
; Sequence 75255, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 75255
; LENGTH: 1643
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_168664C.1
US-10-425-115-75255

Query Match   7.1%; Score 50.2; DB 20; Length 1643;
Best Local Similarity   59.4%; Pred. No. 0.00014;
Matches   85; Conservative   0; Mismatches 58; Indels   0; Gaps   0;

```
Qy      8 CCCTGGATCGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTTAG 67
            |||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    325 CCTTCGGCCAAGTGCGCGTCCTCATCGTCGGTGATTCAGGTGTTGGGAAATCTTCATTGG 384

Qy     68 TCCATCTCCTATGCCAAAATCAAGTGCTGGGAAATCCATCATGACTGTGGGCTGCTCAG 127
```

```
Db    385 TGCATCTCATTTTGAAAAGTTCTGCCATTGCTCGACCATCTCAAACAGTAGGATGCACTG 444
              |||||||| ||  ||    |||||       |||||| ||||
Qy    128 TGGATGTCAGAGTTCATGATTAC 150
              ||| |||    ||||
Db    445 TGGGCATTAAACATGTTACTTAC 467
```

RESULT 13
US-10-437-963-650
; Sequence 650, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
;   APPLICANT: Zhou, Yihua
;   APPLICANT: Cao, Yongwei
;   APPLICANT: Wu, Wei
;   APPLICANT: Boukharov, Andrey A.
;   APPLICANT: Barbazuk, Brad
;   APPLICANT: Li, Ping
;   TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53221)B
;   CURRENT APPLICATION NUMBER: US/10/437,963
;   CURRENT FILING DATE: 2003-05-14
;   NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 650
;   LENGTH: 761
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(761)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_100593C.1
US-10-437-963-650

Query Match   6.4%; Score 45.2; DB 19; Length 761;
Best Local Similarity 50.0%; Pred. No. 0.0036;
Matches 175; Conservative 0; Mismatches 163; Indels 12; Gaps 2;

```
Qy      2 TGGCGTCCCTGGACTCGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTT  61
            |||||  |||   |||  |||||||||||||||||||  |||||||||||||||||||||
Db    122 TGGGCCCCTTCGGCCAAGTGCGCGTCCTCGTCGTCGCGCCGATTCAGGTGTTGGGAAATCCT 181

Qy     62 CGTTAGTCCATCCATCCTATGCCAAAATCAAGTGCTGGGAAATCATCATGACTGCTGGGCT 121
            || ||  |||  ||   ||||  ||||||||||||||  ||||||| ||||||  ||||
Db    182 CTTTGGTGCATCTCATTTTAAAGGGTTCTGCTATTGCTCGACCTGCCCAGACAGTAGGAT 241

Qy    122 GCTCAGTGGATGTCAGAGTTCATGATTACAAAGAAGGAACCCCAGAAGAGAAGAGACCTACT 181
            ||||||||||||||||||||||||||||   |||| ||| |||||||||||||  ||  ||
Db    242 GTGCGGGTTGGCGTTAAACATATTACTTATGATGAAGTGAAGTGCAGGCGGGTTCTTCTAATAACATCA 301

Qy    182 ACATAGAATTATGGGA--------TGTTGGAGGCTCTGTGGGCAGTGCCAGCA-------GCG 229
            |||| ||   ||||||        ||||| |||| |||||||||  ||||||||
Db    302 TCAGCGACGTTCAAAGGAACTTCTTTGTTGAGCTTTGGGGATGTTTCAGGACATGAACGCT 361

Qy    230 TGAAAAGCACAAGAGCACGTATTCCTACAACTCCGTAAATGATTATTATTTTCGTACACGACT 289
            |||||  |||||||||||  |||| |||||  |||||| ||||||||  |||||||||||
Db    362 ACAGAACATGCCGTTCAATTTTCTATACACAAAATTAATGGTGTCATATTTGTTTATGACC 421

Qy    290 TAACAAATAAGAAGTCCTCCCAAAACTTGCGTCGTTGGTCATTGGAAGCT 339
            |||||||    ||   ||||||||||  ||| |||||||  |||||
Db    422 TCTCTCAGAGGAAGACCAAAACAAATTTGAATAAATGGGCAGTTGAAGTT 471
```

RESULT 14
US-10-437-963-25042/c
; Sequence 25042, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa, Thomas J.
;   APPLICANT: Kovalic, David K.
;   APPLICANT: Zhou, Yihua
;   APPLICANT: Cao, Yongwei
;   APPLICANT: Wu, Wei
;   APPLICANT: Boukharov, Andrey A.
;   APPLICANT: Barbazuk, Brad
;   APPLICANT: Li, Ping
;   TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53221)B
;   CURRENT APPLICATION NUMBER: US/10/437,963
;   CURRENT FILING DATE: 2003-05-14
;   NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 25042
;   LENGTH: 1584
;   TYPE: DNA
;   ORGANISM: Oryza sativa
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_29969C.1
US-10-437-963-25042

Query Match   6.3%; Score 44.6; DB 19; Length 1584;
Best Local Similarity 57.6%; Pred. No. 0.0082;
Matches   80; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy     16 CGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTTAGTCCATCTC  75
            | |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1223 CAGGTGCGCGTGCTCCGTCGTCGGCGACTCAGGTGTGGGGAAATCATCATTGGTTCATCTC 1164

Qy     76 CTATGCCAAATCAAGTGCTGGGAAATCATCATGACTGCTGGGCT  135
            ||   |||||||||||||||| |||||   |||||||||| |||||
Db   1163 ATTCTGAAAGGCTCGCAATTGCTGCCACCACCCCCAAAACGATTGGATGCGCAGTTGATGTT 1104

Qy    136 AGAGTTCATGATTACAAAG 154
            ||    ||||  |||||||
Db   1103 AAATATATTTTATTGAAG 1085
```

RESULT 15
US-10-424-599-109350
; Sequence 109350, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa Thomas J
;   APPLICANT: Kovalic David K
;   APPLICANT: Zhou Yihua
;   APPLICANT: Cao Yongwei
;   TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53223)B
;   CURRENT APPLICATION NUMBER: US/10/424,599
;   CURRENT FILING DATE: 2003-04-28
;   NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 109350
;   LENGTH: 610
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(610)
;   OTHER INFORMATION: unsure at all n locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_69758C.1
US-10-424-599-109350

Query Match   5.3%; Score 38; DB 18; Length 610;
Best Local Similarity 55.2%; Pred. No. 0.6;
Matches   74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
Qy     16 CGGGTGAAGGTACTGGTGTTGGGAGACTCAGGTGTTGGGAAATCTTCGTTAGTCCATCTC  75
            | |||||  || |||  |||||||||||||||||||||||||||||||||||||||||||
Db    353 CAGGTCAGAGTCCTCGTTGTTGGTGTGATTCAGGTGTGGGGTTCAGGTGTTCGGGAAGACTTCTTTAGTTAACCTG 412
```

```
Qy    76 CTATGCCAAATCAAGTGCTCGGGAAATCCATCATGGACTGTGGGCTGCTCAGTGGATGTC 135
             |   |  |||    |  |||| ||| |||| |||||||| ||||  || |||||
Db   413 ATTGTAAAAGGTTCTTCAATTGCTCGCCCTCCCTCAAACAATTGGTTGTTCAGTTGATGTG 472

Qy   136 AGAGTTCATGATTA 149
            |  | | |||
Db   473 AAGCATATTACTTA 486
```

Search completed: June 30, 2005, 23:57:30
Job time : 495.231 secs

THIS PAGE BLANK (USPTO)